UCLA
UCLA Electronic Theses and Dissertations

Title
Genomic and metagenomic studies of the gut and skin microbiome with probiotic applications

Permalink
https://escholarship.org/uc/item/79x5d64c

Author
Liu, Jared

Publication Date
2016

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Genomic and metagenomic studies of the gut and skin microbiome with probiotic applications

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Molecular and Medical Pharmacology

by

Jared Liu

2016

© Copyright by

Jared Liu

2016

ABSTRACT OF THE DISSERTATION

Genomic and metagenomic studies of the gut and skin microbiome with probiotic applications

by

Jared Liu

Doctor of Philosophy in Molecular and Medical Pharmacology

University of California, Los Angeles, 2016

Professor Huiying Li, Chair

Growing knowledge of the human microbiome has increased interest in probiotics, which can affect both the host and the microbiota. Genomic and metagenomic analyses can elucidate these effects and thereby provide important considerations for probiotic therapy. This dissertation describes genomic and metagenomic studies of two potential probiotic microorganisms, the gut bacterium *Lactobacillus johnsonii* and the bacteriophage of the skin commensal *Propionibacterium acnes*.

Studies in an Atm-deficient mouse model of ataxia telangiectasia revealed that lymphoma latency, lifespan, and systemic genotoxicity are associated with the abundances of specific intestinal microbes, such as *L. johnsonii*. We isolated a strain of this species, 456, and observed that systemic genotoxicity and inflammation were reduced when 456 was inoculated into these mice. Strain 456 also reduced genotoxicity in wild-type mice but exacerbated genotoxicity induced by whole-body proton irradiation. Genome comparison of strain 456 with 8 other sequenced *L. johnsonii* strains revealed that 456 was genetically distant from other mouse

isolates. Analysis of non-core genomic regions additionally revealed potential host-attachment proteins specific to the genome of 456.

I next investigated the diversity and host interactions of phages that infect *P. acnes*, a commensal bacterium implicated in acne pathogenesis and several other diseases. Genome comparison of *P. acnes* phages isolated from acne patients and healthy individuals revealed that populations of these viruses are often dominated by one strain and that phage strains could be shared among both related and unrelated individuals. Phage infection can result in lysis, pseudolysogeny, or resistance depending on the *P. acnes* lineage. Metagenomic analysis further revealed that phage populations are prevalent at varying abundance across individuals, are highly personal, and show limited genetic change over time within individual strains. *P. acnes* phages may potentially be transmitted between related individuals.

Based on these studies, the effectiveness of *L. johnsonii* probiotics may depend on strain genetics as well as the conditions under which they are applied. Effective phage therapy for *P. acnes*-associated diseases may depend on the structure and dynamics of an individual's *P. acnes* and phage populations. These considerations merit further investigation in the context of these microbes as well as other probiotics.

The dissertation of Jared Liu is approved.

Jing Huang

Jonathan Braun

Ren Sun

Robert Schiestl

Huiying Li, Committee Chair

University of California, Los Angeles

2016

TABLE OF CONTENTS

ABSTRACT OF THE DISSERTATION ii

Acknowledgements vii

Vita x

Chapter 1: Introduction 1

Chapter 2: Intestinal Bacteria Modify Lymphoma Incidence and Latency by Affecting Systemic Inflammatory State, Oxidative Stress, and Leukocyte Genotoxicity 5

Abstract 6

Introduction 6

Materials and Methods 7

Results 8

Discussion 13

Chapter 3: Gut microflora and probiotic treatment affect radiation-induced genotoxicity 17

Abstract 18

Introduction 18

Methods 19

Results 22

Discussion 25

Chapter 4: Genome sequencing and comparative analysis of *Lactobacillus johnsonii* 456 . 27

Abstract 28

Introduction 28

Methods 29

Results 30

Discussion 37

Chapter 5: The diversity and host interactions of *Propionibacterium acnes* bacteriophages on human skin 39

Abstract 40

Introduction 40

Materials and Methods 41

Results 43

Discussion 50

References 53

Chapter 6: Metagenomic analysis of *Propionibacterium acnes* bacteriophage populations in the human skin microbiome 56

Abstract 57

Introduction 57

Methods 58

Results 63

Discussion 73

Chapter 7: Conclusions 75

Divergent modulation of systemic genotoxicity by *L. johnsonii* and the gut microbiota: treatment context may matter 76

Comparative genome analysis can provide insights into probiotic function 76

Diversity and host interactions of *P. acnes* bacteriophages: microbiome structure may also matter 77

Phage dynamics: temporal stability and transmissibility of probiotic treatment 77

Appendix 79

References 82

Acknowledgements

I would like to thank my mentors Drs. Huiying Li and Robert Schiestl, not only for training me in the many aspects of scientific investigation, but, more importantly, for teaching me not to give up, especially during difficult times in my program. I also want to thank my committee members, Drs. Jing Huang, Jonathan Braun, and Ren Sun for helping me navigate through the obstacles and uncertainties of my graduate program.

I want to thank Riceley Yan, Qiao Zhong, and Sam Ngo in the Li Lab for their hard work on testing each of the *P. acnes* strains against each of the phages, as well as Drs. Emma Barnard and Baochen Shi for their work on the metagenomic sequencing and analysis. In the Schiestl Lab, I want to thank Dr. Lynn Yamamoto for helping me get started with the *L. johnsonii* experiments, as well as Drs. Aya Westbrook and Aaron Chapman for teaching me various genotoxicity assays and how to work with mice. I would also like to acknowledge Dr. Irene Maier, who I collaborated with to test the effects of *L. johnsonii* strain 456 on wild-type and irradiated mice described in Chapter 3.

I want to thank my other labmates from the Li and Schiestl Labs: Emily Curd, Nathanael Bangayan, Tremylla Johnson, and Drs. Shi Feng, Mizue Naito, Shuta Tomida, and Sorel Fitz-Gibbon from the Li lab, as well as Michael Davoren, Daniel Malkin, and Drs. Liuba Parfenova and Zorica Scuric from the Schiestl lab, for their helpful suggestions and input during lab meetings and for keeping my life in lab colorful and exciting. Similarly, I also want to thank my classmates from the ACCESS program, especially my roommates Joseph Argus and Dr. Elvira Khialeeva, for being great friends and for making my life outside of lab equally colorful and exciting.

I want to thank my parents and my brother both for being supportive of my decision to go to graduate school and for helping me stay reasonably sane and motivated during stressful times in my research. I also want to thank the Department of Molecular and Medical Pharmacology for giving me the opportunity to do this research, to share it with other people, and also have a good time doing it.

Lastly, I would like to acknowledge funding from NIH grants R01GM099530 and UH2AR057503 from NIGMS and NIAMS as well as NASA grant NNX11AB44G and NIH grant ES09519 that supported the work I have done in the Li Lab and the Schiestl Lab. I also want to acknowledge funding support from the Microbial Pathogenesis Training Grant T32AI07323.

Chapter 2 is adapted from "Yamamoto ML, Maier I, Dang AT, Berry D, Liu J, Ruegger PM, Yang J-I, Soto P a, Presley LL, Reliene R, Westbrook AM, Wei B, Loy A, Chang C, Braun J, Borneman J, Schiestl RH. 2013. Intestinal Bacteria Modify Lymphoma Incidence and Latency by Affecting Systemic Inflammatory State, Oxidative Stress, and Leukocyte Genotoxicity. Cancer Res 73(14):4222–4232".

Chapter 3 is adapted from "Maier I, Liu J, Ruegger PM, Borneman J, and Schiestl RH. Intestinal bacterial phylotype indicators convey radiation-induced genotoxicity in microbiota-restricted and conventional mice" (in preparation).

Chapter 5 is adapted from "Liu J, Yan R, Zhong Q, Ngo S, Bangayan NJ, Nguyen L, Lui T, Liu M, Erfe MC, Craft N, Tomida S, Li H. 2015. The diversity and host interactions of Propionibacterium acnes bacteriophages on human skin. ISME J 9(9):2078–2093".

The appendix chapter is adapted from “Liu J, Cheng A, Bangayan NJ, Barnard E, Curd E, Craft N, Li H. 2014. Draft Genome Sequences of Propionibacterium acnes Type Strain ATCC6919 and Antibiotic-Resistant Strain HL411PA1. Genome Announc 2(4):e00740-14”.

Vita

Education

2004 – 2008 University of California, Berkeley, BA in Molecular and Cell Biology

Teaching

2011 Teaching Assistant for MIMG 109AL, University of California, Los Angeles

2013 Teaching Assistant for MIMG 109AL, University of California, Los Angeles

Awards and Traineeships

2012 Best Group Poster Award, 19th International Microbial Genomics Conference, Lake Arrowhead, California

2012 Microbial Pathogenesis Training Grant, University of California, Los Angeles

2016 Microbial Pathogenesis Training Grant, University of California, Los Angeles

Publications

Liu J, Yan R, Zhong Q, Ngo S, Bangayan NJ, Nguyen L, Lui T, Liu M, Erfe MC, Craft N, Tomida S, Li H. 2015. The diversity and host interactions of *Propionibacterium acnes* bacteriophages on human skin. ISME Journal 9(9):2078–2093. PMID: 25848871.

Liu J, Cheng A, Bangayan NJ, Barnard E, Curd E, Craft N, Li H. Draft Genome Sequences of *Propionibacterium acnes* Type Strain ATCC6919 and Antibiotic-Resistant Strain HL411PA1. 2014. Genome Announcements 2(4):e00740-14. PMID: 25125638

Tomida S, Nguyen L, Chiu BH, **Liu J**, Sodergren E, Weinstock GM, Li H. 2013. Pan-genome and comparative genome analyses of *Propionibacterium acnes* reveal its genomic diversity in the healthy and diseased human skin microbiome. mBio 4(3):e00003-13 PMID: 23631911.

Yamamoto ML, Maier I, Dang AT, Berry D, **Liu J**, Ruegger PM, Yang JI, Soto PA, Presley LL, Reliene R, Westbrook AM, Wei B, Loy A, Chang C, Braun J, Borneman J, Schiestl RH. 2013. Intestinal bacteria modify lymphoma incidence and latency by affecting systemic inflammatory state, oxidative stress, and leukocyte genotoxicity. Cancer Research 73(14):4222-32 PMID: 23860718.

Chapter 1

Introduction

The past decade of research on the human microbiome has expanded perceptions about the relationship between humans and their associated microbes. Building upon the past two centuries of work that linked specific microbial pathogens to many infectious diseases, a new era of investigation is now underway to understand the collective impact of entire microbial communities on human health and disease. Aided by advances in sequencing technologies, studies of the microbiota inhabiting various human body sites have revealed previously unconsidered connections between many inflammatory and metabolic diseases (1, 2), as well as the development of certain cancers and neurological and psychological disorders (3), to disturbances in the relevant microbial communities. Combined with further evidence that these altered communities produce specific metabolites, signaling factors, or physiological alterations relevant to these diseases (1), these observations established the current understanding that microbial dysbiosis, even in the absence of a clear pathogen, can play an important role within the multifactorial etiology of these diseases.

Growing knowledge of the health impacts of microbial communities have naturally led to increased interest and changing perspectives on the use of probiotics, which are defined, viable microorganisms that are administered to specific sites, typically the gut, to treat or prevent disease. Well-documented effects of probiotics include their ability to directly affect the host via immunomodulatory or metabolic mechanisms (4). However, many of their purported health benefits still lack experimental confirmation, and their efficacy for some diseases remains questionable in light of conflicting studies (4). As probiotics have more recently been shown to alter the composition of the endogenous flora in humans (5) and in animal models (4, 6, 7), investigation of their community-level effects may permit a clearer evaluation of their effectiveness.

Genomic and metagenomic analyses can elucidate these host- and microbiota-directed effects of a potential probiotic microbe and thereby provide important considerations for its therapeutic application. For instance, sequencing and comparison of genomes from probiotic and related but non-probiotic strains can provide insight into the mechanisms of host-specific adaptations (8) and probiotic effects (9). Genetic comparisons between probiotic strains as well as investigations of how their abundance interacts with host genetics have also been used to form hypotheses about the interpersonal viability and compatibility of probiotic strains (10, 11). Finally, metagenomic analyses have demonstrated that probiotics can change the composition and diversity of the gut microbiota (12, 13) and that these alterations may be site-specific within the target organ (14).

This dissertation describes genomic and metagenomic studies of two potential probiotic microorganisms, the gut bacterium *Lactobacillus johnsonii* and a bacteriophage of the skin bacterium *Propionibacterium acnes*. *L. johnsonii* is a Gram-positive commensal bacterium found in the gut microbiota of many vertebrates, including humans. It is a common commercial probiotic that has been studied for its pathogen-inhibiting (15–17) and immunomodulatory (18, 19) properties. This latter capacity to reduce host inflammation and promote immunotolerance may underlie a broader application of this probiotic, and modification of the gut microbiota in general, to the prevention or mitigation of genotoxicity in individuals genetically or environmentally vulnerable to DNA damage. The first study evaluates this hypothesis using mouse models of genotoxicity and a strain of *L. johnsonii*, 456, that has been isolated from these mice. The host-targeted effects of this isolate are investigated through *in vivo* experiments and by comparison of its genome to other sequenced *L. johnsonii* strains.

The second study investigates the microbiota-targeted effects of *P. acnes* bacteriophages, which can potentially be used as a skin probiotic. *P. acnes* is a major human skin commensal and has been implicated in the common skin disease acne vulgaris as well as other infections. Since its discovery in the early 1960s, the *P. acnes* phage had been scantly studied outside of its use as a typing method for *P. acnes* strains. However, there has been increased interest in applying the lytic activity of this phage to treat *P. acnes*-related diseases. Effective therapeutic application of phages to control pathogenic *P. acnes* strains requires an understanding of their ecology and bacterium-phage interactions, which is currently lacking. Thus, genome analysis and comparison were used to characterize the diversity of *P. acnes* phages and their interaction with *P. acnes* in the human skin microbiome. An analysis of skin metagenomes was then performed to determine the inter- and intra-individual diversity and dynamics of phage populations. The resulting insights into the distribution, host-specificity, and population-level changes of these phages can form the theoretical foundation for therapeutic strategies that aim to reduce pathogenic *P. acnes* strains in an individual's skin microbiota with minimal disturbance to commensal or probiotic *P. acnes* strain populations.

Chapter 2

Intestinal Bacteria Modify Lymphoma Incidence and Latency by Affecting Systemic Inflammatory State, Oxidative Stress, and Leukocyte Genotoxicity

Intestinal Bacteria Modify Lymphoma Incidence and Latency by Affecting Systemic Inflammatory State, Oxidative Stress, and Leukocyte Genotoxicity

Mitsuko L. Yamamoto[1], Irene Maier[2], Angeline Tilly Dang[3], David Berry[6], Jared Liu[1], Paul M. Ruegger[5], Jiue-in Yang[5], Phillip A. Soto[5], Laura L. Presley[5], Ramune Reliene[1], Aya M. Westbrook[1], Bo Wei[1], Alexander Loy[6], Christopher Chang[3,4], Jonathan Braun[1], James Borneman[5], and Robert H. Schiestl[1,2]

Abstract

Ataxia-telangiectasia is a genetic disorder associated with high incidence of B-cell lymphoma. Using an ataxia-telangiectasia mouse model, we compared lymphoma incidence in several isogenic mouse colonies harboring different bacterial communities, finding that intestinal microbiota are a major contributor to disease penetrance and latency, lifespan, molecular oxidative stress, and systemic leukocyte genotoxicity. High-throughput sequence analysis of rRNA genes identified mucosa-associated bacterial phylotypes that were colony-specific. *Lactobacillus johnsonii*, which was deficient in the more cancer-prone mouse colony, was causally tested for its capacity to confer reduced genotoxicity when restored by short-term oral transfer. This intervention decreased systemic genotoxicity, a response associated with reduced basal leukocytes and the cytokine-mediated inflammatory state, and mechanistically linked to the host cell biology of systemic genotoxicity. Our results suggest that intestinal microbiota are a potentially modifiable trait for translational intervention in individuals at risk for B-cell lymphoma, or for other diseases that are driven by genotoxicity or the molecular response to oxidative stress. *Cancer Res; 73(14); 4222–32. ©2013 AACR.*

Introduction

Ataxia-telangiectasia is an autosomal recessive disorder associated with high incidence of lymphoid malignancies, neurological degeneration, immunodeficiency, radiation sensitivity, and genetic instability (1). Approximately 30% to 40% of all patients with ataxia-telangiectasia develop neoplasia during their life (2): more than 40% of tumors are non–Hodgkin B-cell lymphomas, about 20% acute lymphocytic leukemias, and 5% Hodgkin lymphomas (3). Ataxia-telangiectasia is caused by biallelic mutations in the ATM gene. More than 600 different ATM mutations have been described (www.LOVD.nl/ATM). The ATM gene encodes a ~350-kDa protein, phosphoinositide 3-kinase, which is expressed abundantly in multiple tissues (4) and plays an important role in cell-cycle checkpoint control as well as repair responses to DNA double-strand breaks (DSB; ref. 5). Absence of functional ATM protein results in chromosomal breakage and rearrangements, aberrant V(D)J recombination, and heightened sensitivity to radiation and chemicals with radiomimetic and prooxidant activity.

Although investigations into ataxia-telangiectasia have been greatly enhanced by the development of mouse models, disease penetrance in genetically identical mouse colonies at different laboratories can vary widely. Some ATM-deficient ($Atm^{-/-}$) mice develop early lymphomas and have short lifespans (2–5 months; ref. 6), whereas others display dramatically delayed phenotypes, where 50% of the mice remain viable after 7 to 12 months (7, 8). Lifespan studies on inbred and mixed background mice have failed to show phenotypic differences (9), suggesting that other factors besides genetic diversity are contributing to disease penetrance. Environmental factors such as housing conditions and diet have been postulated to be contributing factors (10). In this study, we examined another potential contributor—the intestinal microbiota.

Intestinal bacteria have been implicated in several types of cancer. In animal models of colorectal cancer, lower incidences in germ-free or antibiotic-treated animals point toward intestinal microbes playing a causative role (11, 12). *Helicobacter* species have been associated with enhanced carcinogenesis including liver cancer, colon cancer, and mammary carcinoma (13). Bacterial products have also been associated with

Authors' Affiliations: [1]Department of Pathology and Lab Medicine, David Geffen School of Medicine, [2]Department of Environmental Health Science, Fielding School of Public Health, [3]Department of Microbiology, Immunology, and Molecular Genetics, University of California; [4]Division of Gastroenterology, Cedars-Sinai Medical Center, Los Angeles; [5]Department of Plant Pathology and Microbiology, University of California, Riverside, California; and [6]Department of Microbial Ecology, Faculty of Life Sciences, University of Vienna, Vienna, Austria

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Current address for M.L. Yamamoto: Pfizer Inc., Experimental Medicine, South San Francisco, California 94080; and current address for L.L. Presley: Laboratory of Lymphoid Tissue Development, Institut Pasteur, Paris 75724, France.

Corresponding Authors: James Borneman, University of California, Riverside, 3401 Watkins Dr, Riverside, CA 92521. Phone: 951-827-3584; Fax: 951-827-4294; E-mail: borneman@ucr.edu; and Robert H. Schiestl. Phone: 310-267-2087; Fax: 310-267-2578; E-mail: rschiestl@mednet.ucla.edu

doi: 10.1158/0008-5472.CAN-13-0022

©2013 American Association for Cancer Research.

increased immune activation and inflammation as well as hyperplasia (14, 15). Conversely, probiotic formulations containing lactic acid bacteria have been shown to reduce the incidence of chemically mediated hepatocellular carcinoma and colon cancer in rats (16).

The objective of this study was to examine the role of intestinal bacteria in the penetrance of lymphoma in $Atm^{-/-}$ mice. We first show the effects of housing and intestinal bacteria on lifespan, lymphoma latency, oxidative stress, and systemic DNA damage (or genotoxicity). Next, high-throughput sequence analysis was used to identify mucosa-associated bacteria from animals reared in 2 distinct housing conditions, differing in intestinal microbiota and the aforementioned ATM-deficient traits. Finally, after determining that *Lactobacillus johnsonii* was higher in abundance in our more cancer-resistant mouse colony, we showed its ability to reduce systemic inflammation and genotoxicity when administered to animals from our more cancer-prone colony.

Materials and Methods

Animal housing and husbandry

$Atm^{-/-}$ mice were obtained by intercrossing $Atm^{+/-}$ mice and identified by genotyping (see Supplementary Materials and Methods for additional details). Mice were housed under standard conditions in accordance with the Animal Research Committee at University of California Los Angeles (UCLA, Los Angeles, CA). Mice were housed under 2 types of specific pathogen-free (SPF) conditions, where either sterile (SPF-S) or nonsterile (SPF-N) food, water, and bedding were used. $Atm^{-/-}$ mice harboring restricted microbiota (RM) and conventional microbiota (CM) were created by rederivation as described in the study by Fujiwara and colleagues (17) and by antibiotic treatment (18) followed by orogastric gavage of CM feces, respectively.

Mouse longevity studies

For the longevity studies, $Atm^{-/-}$ mice were kept until they developed signs of tumors or became sick according to ARC protocols, at which time they were euthanized, or if nonsymptomatic kept until they were found dead. Mice were sent to a veterinary pathologist for necropsy. Differences in longevity and lymphoma latency were analyzed using the log-rank test.

Pun reversion assay

Pun reversions were counted in mice that were 10-days-old. Pun reversions can be seen as a black spot on the fur and were counted as described previously (19). Statistics were done using the χ^2 test.

Oxidative stress

Frozen blood was prepared for enzymatic measurement of the oxidized form of glutathione using the Bioxytech GSH: GSSG-412 Kit Assay (Oxis). Reduced GSH and GSSG were determined separately by reaction with glutathione reductase. The colorimetric assay was conducted in triplicate for each blood sample from all male experimental $Atm^{-/-}$ mice. Data were expressed as the ratio of free GSH to GSSG.

Micronucleus assay

Micronuclei were examined in peripheral blood erythrocytes collected from ~6-month-old mice and stained with Wright-Giemsa (Sigma-Aldrich). At least 2,000 erythrocytes were counted at ×100 magnification as described previously (20). Statistics were done using Student *t* tests.

Alkaline comet assay

DNA strand breaks were measured in peripheral blood cells using the alkaline comet assay. Blood was collected from the facial vein of mice approximately 6-months-old and diluted 1:1 with RPMI + 20% dimethyl sulfoxide (DMSO) for storage at −80°C until the assay was conducted. The comet assay was basically conducted as previously described (20). Statistical analyses were done using Student *t* tests.

Bacteria community analyses of SPF-N and SPF-S mice

Fecal pellets from $Atm^{+/-}$ mice (SPF-N and SPF-S) were collected and immediately snap-frozen in liquid nitrogen and stored at −80°C. Nucleic acids from fecal pellets were purified using a phenol–chloroform extraction with bead-beating and a fragment of the 16S rRNA gene targeting the V6–V9 region of most bacteria was amplified with PCR primers 909F (5′-ACTCAAAKGAATWGACGG-3′) and 1492R (5′-NTACCTTGTTACGACT-3′). The template was amplified and tagged with a sample-specific 8-nt barcode sequence via a 2-step, low cycle number, barcoded PCR protocol. Pyrosequencing was conducted on a GS FLX 454 sequencer at the Norwegian Sequencing Centre. Pyrosequencing reads were quality-filtered using LUCY and clustered into phylotypes at 97% identity using UCLUST. Taxonomic classifications were assigned, and alpha and beta diversity metrics were produced using the QIIME software package (21). The VEGAN package was used for permutational multivariate analysis of variance and the indispecies package was used to identify phylotypes that were indicators for a colony type; identifying indicators involves analysis of occurrence and abundance. To focus on abundant indicators, we considered only indicators that had an elevated relative abundance of at least 1% (e.g., 0.5%–1.5%). See Supplementary Materials and Methods for additional details.

Bacteria community analyses of CM and RM mice

Intestinal mucosa samples of CM and RM mice were obtained as described in the study by Presley and colleagues (22). DNA was extracted from these samples using the PowerSoil DNA Isolation Kit (MO BIO Laboratories), and a 30-second beat-beating step using a Mini-Beadbeater-16 (BioSpec Products). One hundred microliter PCR amplification reactions were conducted in an MJ Research PTC-200 thermal cycler (Bio-Rad Inc.). PCR primers targeted a portion of the 16S and 23S rRNA genes and the hypervariable intergenic region, with the reverse primers including a 12-bp barcode (Supplementary Tables S1a and S1b). PCRs contained: 50 mmol/L Tris (pH 8.3), 500 μg/mL bovine serum albumin, 2.5 mmol/L $MgCl_2$, 250 μmol/L of each deoxynucleotide triphosphate (dNTP), 400 nmol/L of forward PCR primer, 200 nmol/L of each reverse PCR primer, 4 μL of DNA template, and 2.5 units JumpStart *Taq* DNA polymerase (Sigma-Aldrich). Thermal cycling parameters

were 94°C for 5 minutes; 35 cycles of 94°C for 20 seconds, 56°C for 20 seconds, and 72°C for 40 seconds, and followed by 72°C for 5 minutes. PCR products were purified using a MinElute 96 UF PCR Purification Kit (Qiagen). DNA sequencing was conducted using an Illumina HiSeq 2000 (Illumina, Inc.). Clusters were created using a template concentration of 2.5 pmol/L. One hundred base sequencing reads of the 5′-end of the amplicons and 7 base barcode reads were obtained using the sequencing primers listed in Supplementary Table S1c. Demultiplexing, quality control, and operational taxonomic units (OTU) binning were conducted using QIIME (21). OTUs were binned at 97% identity.

L. johnsonii isolation and oral inoculation experiments

L. johnsonii (strain LJ-RS-1) was isolated from RM wild-type mouse feces using Lactobacillus Selection Agar (BD). For the oral inoculation experiments, this bacterium was grown on LB agar supplemented with 2% glucose and 0.05% (wt/vol) cysteine at 37°C under anaerobic conditions. The strain was grown overnight and suspended in PBS. Before inoculation with *L. johnsonii*, CM $Atm^{-/-}$ mice were treated with 1 g/L ampicillin (Sigma-Aldrich), neomycin (Thermo Fisher Scientific), and metronidazole (Baxter) and 500 mg/L vancomycin (Hospira) in their drinking water for 1 week as described previously (18). Then, 10^9 colony-forming units (CFU) of *L. johnsonii* was administrated every other day by orogastric gavage to a group of 8 animals for 4 successive weeks; in addition, the drinking water for same group of animals contained 10^9 CFU/mL of *L. johnsonii*. Fecal population densities of *L. johnsonii* were measured before, during, and after *L. johnsonii* administration, using a previously described sequence-selective quantitative PCR (qPCR) assay (22). After 4 weeks, mice were euthanized using 3% isoflurane and analyzed as described below.

Gene expression by real time-PCR

RNA from peripheral blood mononuclear cells (PBMC) or tissue was collected by phenol–chloroform extraction followed by ethanol precipitation. Pellets were resuspended in nucleotide-free water, and RNA was treated with 200 units of DNase I (Ambion) for 2 hours at 37°C. RNA levels of TGF-β, interleukin (IL)-1β, IL-4, IL-10, IL-10, IFN-γ, and myeloid differentiation primary response gene 88 (*MYD88*) were determined from cDNA and compared as relative expression changes in *L. johnsonii*–treated animals compared with PBS-treated animals. Real time (RT)-PCR of gene expression was carried out using the TaqMan Gene Expression Assays System (Applied Biosystems). The level of gene transcription was normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

Flow cytometry

PBMCs or tissue cell pellets were resuspended in staining buffer. Lymphocytes were stained with CD3, CD4, CD19, and CD335 antibodies (Biolegend). Cells were fixed in 4% paraformaldehyde-PBS and scanned on a Becton Dickinson FACS Calibur flow cytometer. A minimum of 25,000 events were collected. Flow cytometry data were analyzed using Tree Star FlowJo software. Statistical analyses (Student *t* test) were used to determine if mean values were different ($P < 0.05$).

Sequence data

DNA sequence data have been deposited at the NCBI Short Read Archive under accession numbers SRA059288 and SRX256360.

Results

Housing affects genetic instability, lifespan, and lymphoma latency of $Atm^{-/-}$ mice

When the Schiestl laboratory moved their mice from Harvard University to UCLA in the year 2000, the median lifespan of their $Atm^{-/-}$ mice began to increase over a 7-year period from approximately 4 to 12 months (unpublished observations). To examine the influence of housing conditions on traits associated with ATM deficiency, we first compared $Atm^{-/-}$ and wild-type mice from 3 housing conditions: (i) Harvard housing and (ii and iii) UCLA housing in SPF conditions with either sterile (SPF-S) or nonsterile (SPF-N) food, water, and bedding. Genetic instability was assessed using the p^{un} reversion assay, which measures DNA deletion events repaired by homologous recombination (19). $Atm^{-/-}$ mice at Harvard had a significantly higher level of genetic instability compared with their wild-type littermates (Fig. 1A; ref. 19). As lifespan increased in the $Atm^{-/-}$ mice over 7 years under SPF-S conditions, DNA deletion frequency decreased to 10%. These levels are comparable to the wild-type mice (Fig. 1A), indicating loss of the abnormally high genetic instability that has been recognized as a hallmark of $Atm^{-/-}$ mice and human cells from ataxia-telangiectasia subjects (19, 23). Conversely, DNA deletion frequencies of $Atm^{-/-}$ mice housed under SPF-N conditions were 43%, a level even higher than those observed in the Harvard colony (Fig. 1A). $Atm^{-/-}$ mice housed under SPF-N conditions also had an increased level ($P < 0.05$) of DNA deletions compared with their wild-type littermates (Fig. 1A).

Focusing on the 2 UCLA $Atm^{-/-}$ mouse colonies, we then compared lifespan and lymphoma latency. SPF-N mice exhibited significantly shorter lifespans than SPF-S mice (44 vs. 51 weeks; Fig. 1B). As $Atm^{-/-}$ mice primarily die from lymphomas (6), we examined the cause of mortality or morbidity. Although the incidence of lymphomas was similar between mice housed in SPF-N (74%) and SPF-S (76.5%) conditions, time of onset was significantly shorter in the SPF-N mice (Fig. 1C). The median age at which SPF-N mice died of lymphoma was 25 weeks compared with 60 weeks for SPF-S mice. These results are consistent with the observed increase in SPF-N genetic instability (Fig. 1A), likely the primary driver of accelerated carcinogenesis. The decreased lymphoma latency in SPF-N mice was also the major cause of the decreased lifespan as the lymphoma-free mice in the SPF-N and SPF-S conditions had similar lifespans (data not shown).

After observing the striking effects of housing on ATM-deficient traits, we examined the intestinal bacteria from fecal pellets of $Atm^{+/-}$ mice in these 2 UCLA mouse colonies. High throughput sequence analysis of bacterial 16S rRNA genes revealed that these 2 colonies harbor distinct microbial communities (Fig. 1D). An indicator analysis identified several species-level phylotypes whose populations were higher in CM versus RM mice, suggesting a possible causative role in

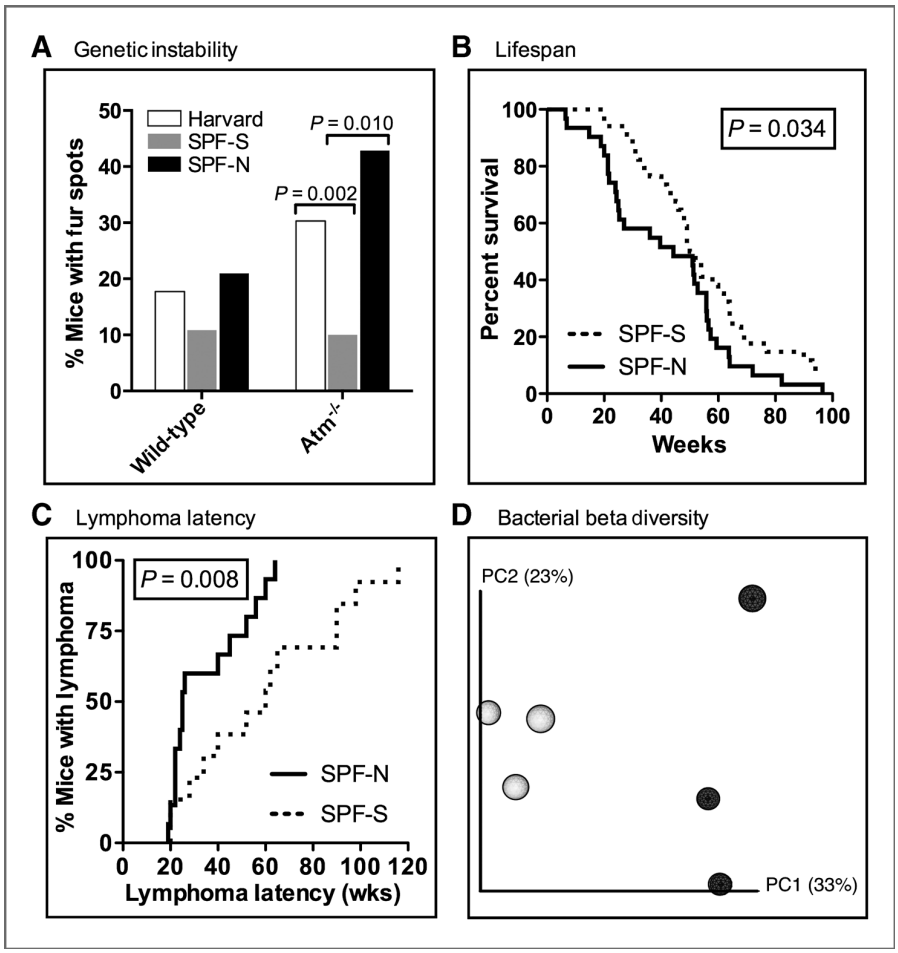


Figure 1. Genetic instability, lymphoma latency, and lifespan are increased in separate isogenic colonies of $Atm^{-/-}$ mice. A, genetic instability (DNA deletions) was measured by P^{un} reversion assays ($n = 22$, 83, and 62 for Harvard, SPF-S, and SPF-N, respectively). B, Kaplan–Meier survival curve ($n = 34$ and 31 for SPF-S and SPF-N, respectively). C, lymphoma latency was determined for the subset of mice developing lymphoma ($n = 13$ and 15 for SPF-S and SPF-N, respectively). D, an unweighted UniFrac analysis of bacteria in $Atm^{+/-}$ mice in SPF-S (light spheres) and SPF-N (dark spheres) conditions. Spheres around points (individual mice) indicate the 95% confidence ranges estimated by bootstrap resampling of data sets at 1,300 reads.

carcinogenesis. One correlate is a member of the Helicobacteriaceae, which are bacteria known to promote cancer (Supplementary Fig. S1).

$Atm^{-/-}$ mice with RM have an increased lifespan and decreased systemic genotoxicity and oxidative stress

To further investigate the role of intestinal microbiota in lymphoma penetrance, we created $Atm^{-/-}$ mice harboring RM and CM, which were created by rederivation as described by Fujiwara and colleagues (17) and by antibiotic treatment (18) followed by orogastric gavage of CM feces, respectively. These models were chosen because CM and RM harbor distinct intestinal microbiota (17, 22, 24), and RM mice possess immunologic traits that can potentially influence lymphoma penetrance, including unusually high levels of cytolytic central memory CD8+ T cells that target neoplastic cells. In addition, as RM mice are physically isolated from other colonies, their immunologic traits have persisted for many years, providing a stable platform for our investigations.

Lifespan in the 2 mouse colonies was significantly different, with median survival of CM mice (31.1 weeks) being shorter than RM mice (45.6 weeks; Fig. 2A). Examination of systemic genotoxicity using different metrics showed that $Atm^{-/-}$ mice with CM microbiota exhibited higher levels of DNA damage than those with RM microbiota. Similar differences were not found between SPF-N and SPF-S mice. Clastogenic DNA damage, determined by the presence of micronuclei in peripheral blood erythrocytes, was nearly 85% higher in CM versus RM mice (Fig. 2B). DNA strand breaks, measured by the alkaline comet assay, were also higher in CM than in RM mice (Fig. 2C). Levels of DNA strand breaks were determined by the Olive tail moment (Fig. 2C) and confirmed by analysis of percent tail DNA at the 80th percentile (data not shown).

As oxidative stress can be produced by intestinal microbiota both directly and indirectly (25, 26), and oxidative stress is associated with both genotoxicity and carcinogenesis (27), we also assessed the antioxidant state of $Atm^{-/-}$ mice harboring CM and RM microbiota. Glutathione (GSH) is a major cellular antioxidant linked to several diseases including cancer. A metric of oxidative stress can be expressed as the ratio of GSH to GSSG (the dimeric oxidized form) in peripheral blood. We found that $Atm^{-/-}$ mice harboring CM microbiota had higher levels of oxidative stress (lower GSH/GSSG ratios) than RM

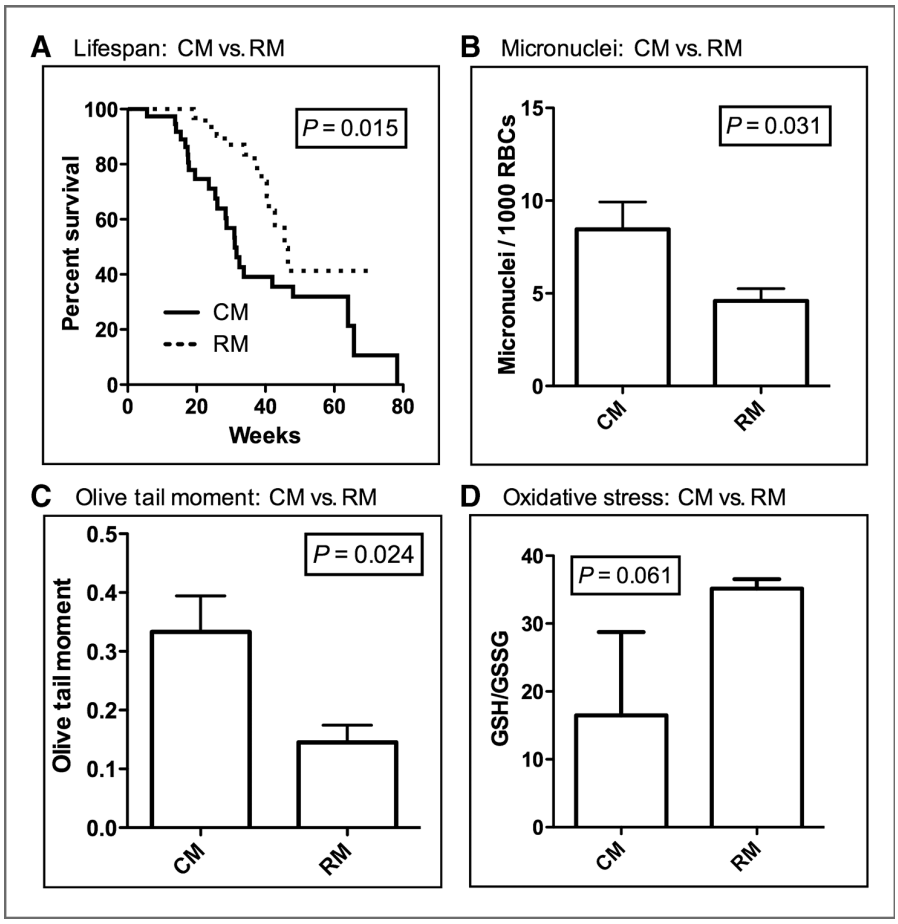


Figure 2. Lifespan, genotoxicity, and oxidative stress of $Atm^{-/-}$ mice bearing CM and RM microbiota. A, Kaplan–Meier survival of CM and RM $Atm^{-/-}$ mice ($n = 38$ and 31 for CM and RM mice, respectively). B, peripheral blood erythrocyte micronuclei levels ($n = 5$ and 6 for CM and RM mice, respectively). C, olive tail moments in blood leukocytes ($n = 5$ for both groups). D, oxidative stress measured by GSH/GSSG ratios ($n = 3$ and 2 for CM and RM mice, respectively). All measurements were in mice ages 4 to 6 months. Error bars indicate SE.

mice (Fig. 2D), suggesting that it may also play a role in lymphoma penetrance in this model system.

The results presented in Fig. 1 indicate that sterility may play a role in carcinogenesis in $Atm^{-/-}$ mice; the results in Fig. 2 suggest that the intestinal microbiota is a major contributor to the phenotypic differences observed within and among isogenic $Atm^{-/-}$ mouse colonies (9).

Microbiota in CM and RM mice are distinct

A high-throughput sequence analysis revealed broad taxonomic differences in bacterial rRNA gene composition from the intestinal mucosa of CM and RM mice. We conducted this analysis because all prior investigations of CM and RM microbiota were relatively shallow in depth (17, 22, 24). A UniFrac analysis showed that the bacterial communities between CM and RM mice are distinct (Fig. 3A and B). The majority of rRNA gene sequences were classified into the phyla Bacteroidetes/Chlorobi (purple, Fig. 3C) and Firmicutes (light green, Fig. 3C). However, the most consistent phyla-level difference between CM and RM was in the Proteobacteria (sage green with asterisks, Fig. 3C), with the most statistically significant differences occurring in the $Atm^{-/-}$ mice (Fig. 3D). Although compositional differences were detected between mouse genotypes ($Atm^{-/-}$ vs. $Atm^{+/+}$) and intestinal regions (small intestine vs. colon), CM versus RM provided the most distinct groupings (Fig. 3A and B). A more comprehensive analysis is provided in Supplementary Table S2.

An analysis of the rRNA gene sequences at a finer taxonomic level revealed several OTUs from the intestinal mucosa that were differentially abundant in CM and RM mice (Fig. 4A). We consider differentially abundant OTUs worthy of further study because they might represent individual bacterial species participating in the enhancement or inhibition of systemic genotoxicity or other pertinent CM-RM metrics. The closest BLAST hit of one of the most abundant OTUs in many RM habitats was *Porphyromonas asaccharolytica* (98% identity to CP002689, 43% coverage; light blue with black asterisks, Fig. 4A). This phylotype was significantly more abundant in RM than CM $Atm^{-/-}$ mice in both the small intestine and colon, and it comprised at least 50% of the total sequencing reads from these habitats (Fig. 4B), warranting further investigation as a potential inhibitor of systemic genotoxicity. Similarly, another OTU exhibiting higher populations in RM habitats of $Atm^{-/-}$ mice (Fig. 4C; and maroon with white asterisks, Fig. 4A) was identical to *L. johnsonii* (100% identity to CP002464, 100% coverage). A phylotype that was more abundant in CM than RM

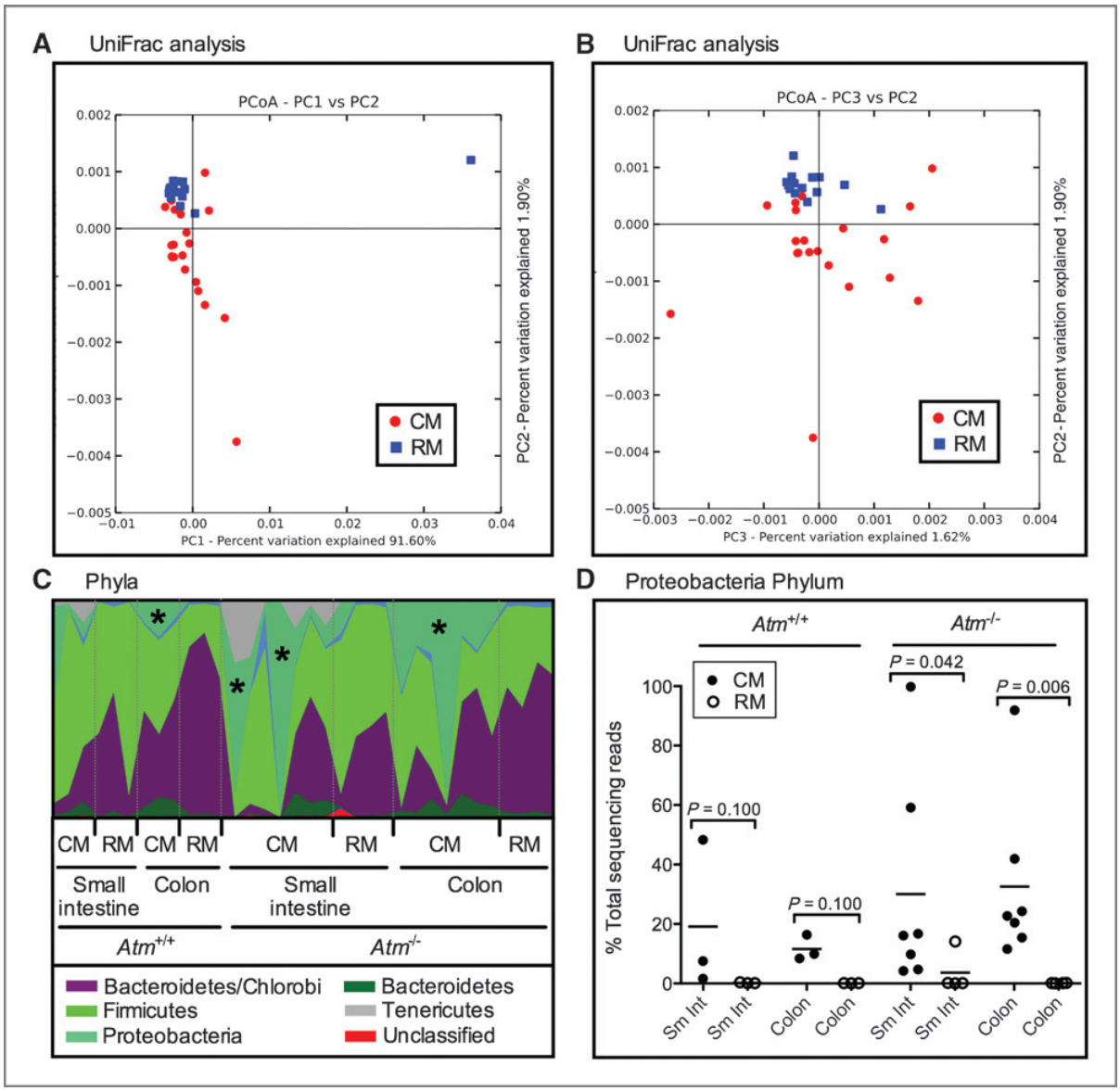


Figure 3. Intestinal microbiota in CM and RM mice are distinct at the community and phyla levels. A and B, a normalized weighted UniFrac analysis of bacteria in mice in CM and RM conditions. C, area plot of phyla in CM and RM mice by genotype and intestinal region. Colored rectangles at bottom designate phyla type. Asterisks in area plot designate Proteobacteria. See Supplementary Fig. S3 for a complete Phyla legend. D, distribution and abundance of Proteobacteria by genotype and intestinal region. For CM, $n = 3, 7, 3, 7$ for small intestine (SI)-$Atm^{+/+}$, SI-$Atm^{-/-}$, colon (CLN)-$Atm^{+/+}$, and CLN-$Atm^{-/-}$, respectively. For RM, $n = 3, 4, 3, 4$ for SI-$Atm^{+/+}$, SI-$Atm^{-/-}$, CLN-$Atm^{+/+}$, and CLN-$Atm^{-/-}$, respectively. Differences were assessed by Mann–Whitney *U* tests.

mice and therefore a candidate for causing the observed genotoxicity, included a member of the Helicobacteriaceae (Supplementary Table S3). A more comprehensive OTU analysis is provided in Supplementary Table S3.

L. johnsonii decreases systemic inflammation and genotoxicity in $Atm^{-/-}$ mice

To verify whether individual bacteria were contributing to the differential systemic genotoxicity detected in CM and RM mice, we carried out oral inoculation experiments. The high-throughput sequence analysis had identified numerous phylotypes exhibiting higher populations in RM than CM mice—a feature suggesting a possible beneficial role. We attempted to culture 2 of the most abundant of these (Fig. 4B and C) and successfully isolated and grew *L. johnsonii* from RM mice in pure culture; the rRNA gene sequence of our *L. johnsonii* isolate is identical to the one identified by the high-throughput sequence analysis (Fig. 4C).

CM $Atm^{-/-}$ mice were orally gavaged with 10^9 CFU of *L. johnsonii* every other day for 4 weeks. Fecal qPCR targeting

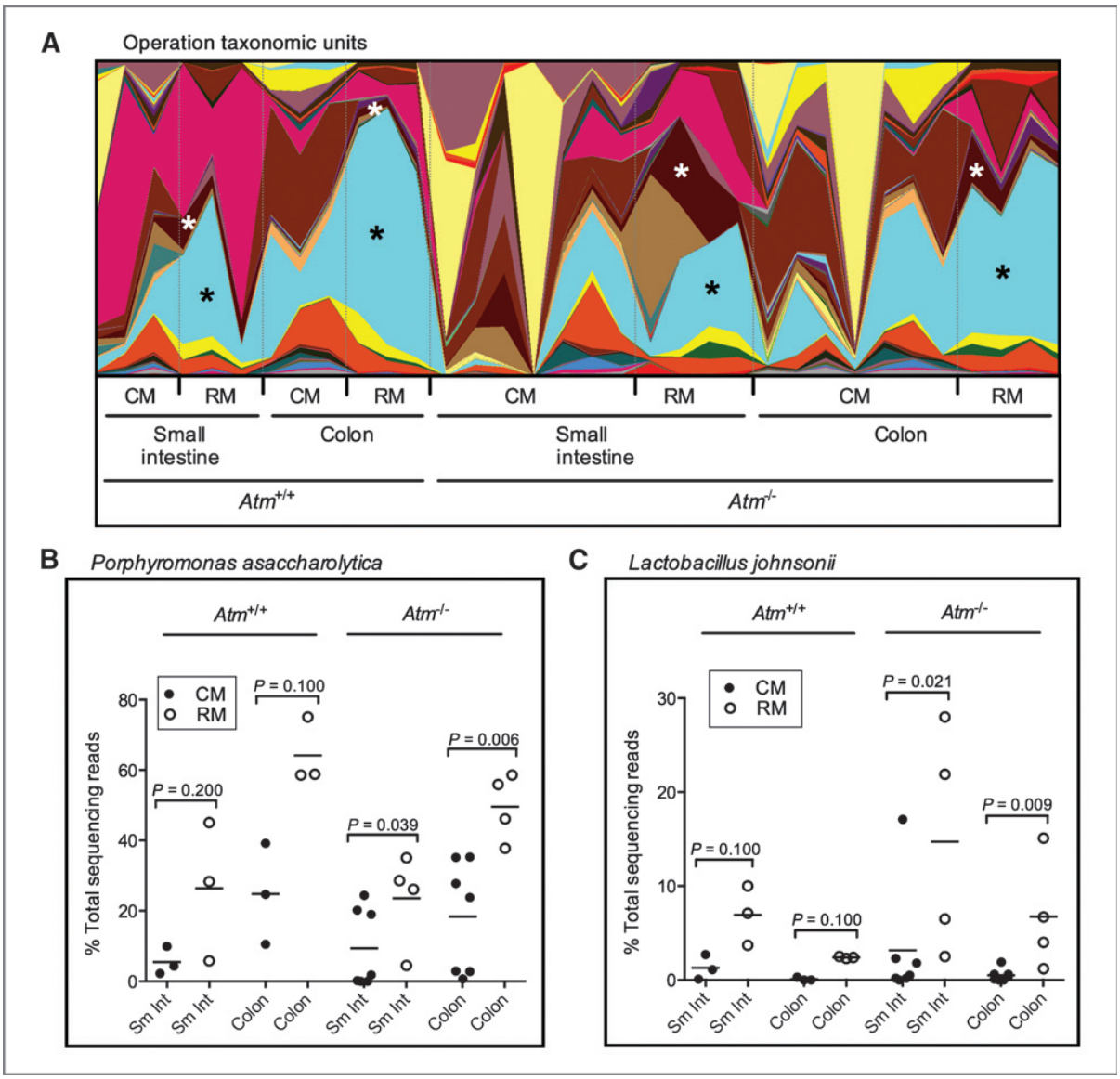


Figure 4. Intestinal microbiota in CM and RM mice are distinct at the OTU level. A, area plot of the most abundant OTUs in CM and RM mice by genotype and intestinal region. Black and white asterisks designate *P. asaccharolytica* and *L. johnsonii*, respectively. B, distribution and abundance of *Porphyromonas asaccharolytica* by genotype and intestinal region. Differences were assessed by Mann–Whitney *U* tests. C, distribution and abundance of *L. johnsonii* by genotype and intestinal region. Differences were assessed by Mann–Whitney *U* tests. See Fig. 3 for mouse numbers per genotype and intestinal region.

L. johnsonii showed again the deficiency of this phylotype in CM $Atm^{-/-}$ mice and also that periodic administration resulted in successful establishment and maintenance of high enteric levels (Fig. 5A). After this 4-week period, *L. johnsonii* but not the vehicle control (PBS) resulted in reduced micronucleus levels (Fig. 5B); however levels of DNA strand breaks were not significantly changed as measured by the COMET assay or γH2AX focus formation (data not shown). This was a time-dependent effect, as no difference was observed after only 1 or 2 weeks of treatment (Supplementary Fig. S2a).

As systemic genotoxicity is induced by innate inflammatory mediators (20, 28), we analyzed the effect of *L. johnsonii* administration on basal systemic inflammatory parameters in these mice. In the liver, *L. johnsonii* significantly reduced the abundance of both hepatic natural killer (NK) and T cells (Fig. 5C and D). A comparable reduction of these leukocyte subsets was also observed in the splenic and blood compartments (Supplementary Fig. S2b and S2c). With respect to molecular mediators, *L. johnsonii* treatment significantly reduced levels of the pro-inflammatory cytokines IL-1β and IFN-γ, and elevated the levels of the anti-inflammatory cytokines TGF-β and IL-10

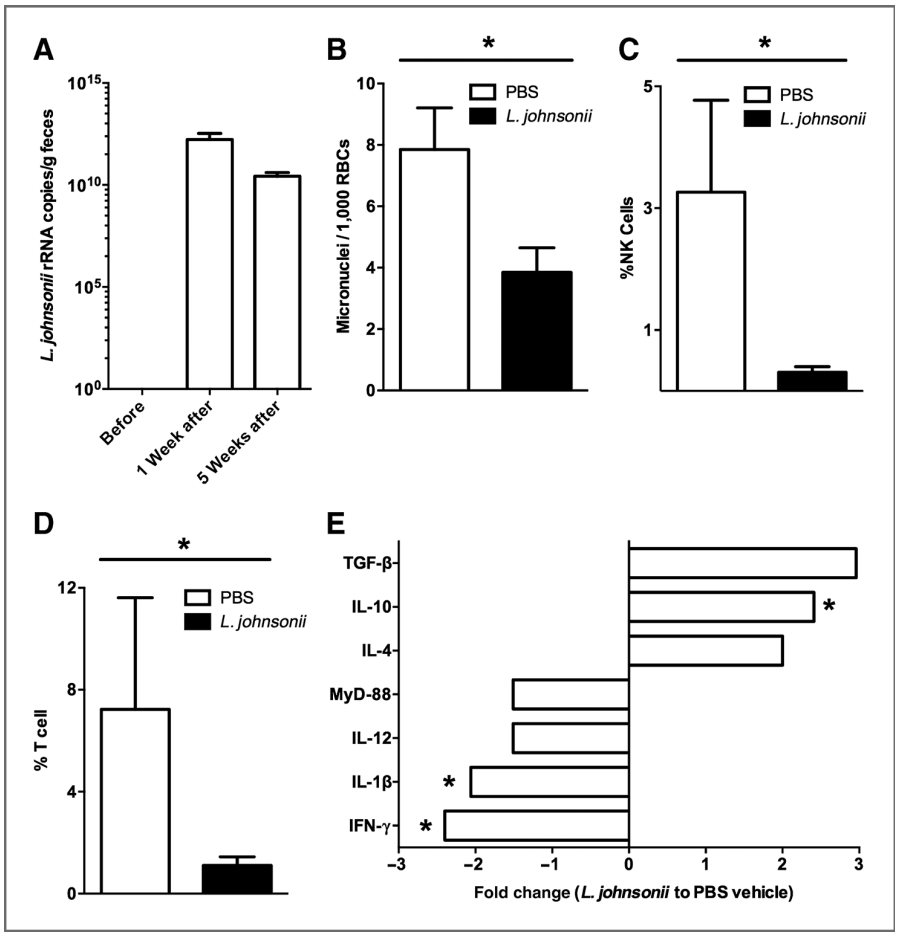


Figure 5. Effect of oral administration of *L. johnsonii* on genotoxicity and inflammation in CM $Atm^{-/-}$ mice. A, fecal levels of *L. johnsonii* were quantified by qPCR. B, micronucleus quantification in RBCs. C and D, levels of hepatic NK (C) and T cells (D) were quantitated by flow cytometry for CD335 and CD3 cells, respectively. E, hepatic tissue levels of cytokines were measured by ELISA, and fold change was calculated for *L. johnsonii* relative to vehicle (PBS) control groups. Error bars indicate SE. *, $P < 0.05$ by Student *t* tests.

(Fig. 5E). Similar changes were also observed in the blood compartment (Supplementary Fig. S2d).

These findings indicate that short-term administration of the single RM-associated bacterium, *L. johnsonii*, recapitulated the reduction of systemic micronucleus formation observed in RM mice, a host effect associated with reduced systemic inflammatory activity.

Discussion

Our results are the first to show a relationship between intestinal microbiota and lymphoma onset. In addition, these investigations generated a detailed catalog of bacterial phylotypes that are differentially abundant between CM and RM mice, thereby providing candidates that may influence a wide range of traits from systemic genotoxicity (this study), oxidative stress (this study), colitis resistance (29), pathogen clearance (30) and selective reduction of marginal zone (MZ) B cells (31), plasmacytoid dendritic cells (pDC; ref. 17), and invariant natural killer (iNK) T cells (24). Moreover, we isolated one highly enriched species, *L. johnsonii*, from RM mice and subsequently showed its ability to decrease systemic inflammation and micronucleus formation in $Atm^{-/-}$ mice via oral inoculation experiments. Below, we propose several mechanistic hypotheses of how *L. johnsonii* might influence these important host traits.

First, *L. johnsonii* may reduce systemic genotoxicity by inhibition of basal intestinal inflammatory activity and its systemic sequelae. Our recent work revealed that intestinal inflammation-associated genotoxicity occurs not only locally but also systemically. Using either a model chemical inflammatory agent, dextran sodium sulfate, or immune-mediated genetic models to induce local intestinal inflammation, we found that systemic genotoxicity was elevated in peripheral lymphocytes, an effect amplified in $Atm^{-/-}$ versus wild-type mice (20). Moreover, such lymphocyte genotoxicity is particularly abundant in the B lymphocyte subset, the progenitor cell type for B-cell lymphoma (the predominant cancer in patients with ataxia-telangiectasia; ref. 32). Genetic and intervention studies have revealed that one factor is systemic dispersal of the intestinal cytokine TNFα, which permits genotoxicity in TNFα receptor–bearing lymphocytes remote from the intestinal compartment, in a process dependent on cell-autonomous, NF-κB and AP-1 induction of reactive oxygen and nitrogen species (RONS; ref. 28).

The present study documents the reduction of several systemic measures of inflammatory activity after *L. johnsonii*

restoration in CM $Atm^{-/-}$ mice. There are several known anti-inflammatory mechanisms of *Lactobacillus* spp. that may account for this host response, including products that directly modulate the NF-κB inflammatory program of mucosal epithelial and hematopoietic cell types, or indirectly by ecologically changing the composition or functional activity of enteric microbial community (33, 34). Our study also documented that enteric *L. johnsonii* resulted in a reduction in the abundance of hepatic and migratory (blood, splenic) NK and T cells; a reduction also concordant with their reduced levels in CM colonies (17, 31). The control of NK and T cells in these compartments is complex, integrating a diversity of chemoattractant and trophic cytokines. *Lactobacillus* spp. may either augment or reduce the production of these key cytokines, by parenchymal epithelial and hematopoietic cells, depending on differences in host compartment and *Lactobacillus* spp. (34, 35). The present study suggests that enteric (dietary) administration of *L. johnsonii* may represent a viable strategy to reduce inflammation-induced genotoxicity.

The ability of *L. johnsonii* to reduce pathogen-associated inflammation has been demonstrated in several prior *in vivo* investigations, including 2 involving *H. pylori*, a known cancer-promoting bacterium. In experiments examining *H. pylori* infection in C57BL/6 mice, oral administration of *L. johnsonii* over a 3-month period reduced the amounts of lymphocytic and neutrophilic infiltration of the lamina propria as well as proinflammatory chemokines (36). Similarly, *L. johnsonii* reduced both *H. pylori* populations and gastritis in Mongolian gerbils (37). A single inoculation of 1-day-old chickens with *L. johnsonii* inhibited the colonization and persistence of *Clostridium perfringens*, a poultry pathogen that causes necrotic enteritis (38). Finally, prior inoculation of gerbils with *L. johnsonii* prevented a persistent infection by the protozoan parasite, *Giardia intestinalis* (39).

Second, analysis of the literature suggests that *L. johnsonii* reduces immune-mediated oxidative stress and systemic genotoxicity by decreasing NF-κB activation. Linked to cancer and various inflammatory disorders, NF-κB is involved in managing responses to a wide range of potentially deleterious stimuli. Consistent with an agent that downregulates NF-κB activation, *L. johnsonii* and other CM-RM–associated microbiota may affect lymphoma penetrance by altering systemic genotoxicity caused by immune-mediated oxidative stress. Our results and others show that oral administration of *L. johnsonii* reduces both oxidative stress (40, 41) and systemic genotoxicity (this study). Mechanistically, there are precedents that microbial environments modulate cancer formation, in part, through oxidative stress mediated by inflammatory or carcinogenic bacterial metabolites on local epithelial cells (15, 42). For example, inflammation that accompanies *H. pylori* infection, *S. haematobium* infection, or human inflammatory bowel disease is associated with elevated risk of stomach, bladder, or colon cancer, respectively (42). In an experimental system examining vaginal infection of mice with the bacterium *Gardnerella vaginalis*, oral administration of *L. johnsonii* reduced levels of proinflammatory cytokines, oxidative stress (iNOS), and activation of NF-κB (41). An important factor linking inflammation to neoplasia is genotoxicity from inflammation-associated oxidative products, created either in trans (due to local oxidative products of inflammatory cells), and cell autonomously (due to endogenous intracellular oxidative products induced by receptors to TNFα and other cytokines; refs. 43, 44). In disease models, enteric colonization of *H. hepaticus* elicits innate immune activation, which via TNF-α and iNOS induction yields host oxidative products required for neoplasia.

Finally, genomic analysis of *L. johnsonii* revealed several features that may contribute to a superior colonizing ability in the mucosa and to its ability to outcompete pathogens and other proinflammatory organisms (45). Attachment to the host is often a key feature of mucosa-associated microbes. Putative cell surface proteins in *L. johnsonii* have similarity to the mucin-binding protein (MUB) from *L. reuteri* (46). In addition, similarities to Fap1 and GspB from *Streptococcus* species suggest *L. johnsonii* encodes adhesive and fimbrial proteins, respectively (47, 48). *L. johnsonii* also produces a putative cell surface protein with similarity to an IgA protease, which could enable it to avoid a key host defense mechanism. In addition, *L. johnsonii* may inhibit potential microbial competitors by producing the bacteriocin Lactacin F and by increasing Paneth cell numbers, which are a host cell type that produces antimicrobial compounds (49–50).

In sum, these investigations build on prior work linking intestinal microbiota, via control of basal levels of inflammation and oxidative stress, to systemic genotoxicity and carcinogenesis. Given that intestinal microbiota is a potentially modifiable trait, these and related insights hold considerable promise for translational intervention of B-cell lymphoma and other diseases driven by immune-mediated oxidative stress and its resulting systemic genotoxicity. The translational promise of this approach, exemplified by the present study, suggests that simple interventions such as sustaining enteric levels of *L. johnsonii* may favorably shift microbial composition and function to reduce basal levels of genotoxicity in a manner that may reduce cancer risk in susceptible individuals, such as those bearing the ataxia-telangiectasia genotype.

Disclosure of Potential Conflicts of Interest

J. Braun and J. Borneman have ownership interest (including patents) as the co-inventor on provisional patent application by the University of California related to this article. R.H. Schiestl has ownership interest (including patents) in Microbio, Pharma Com. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: L. Yamamoto, I. Maier, L.L. Presley, J. Braun, J. Borneman, R.H. Schiestl

Development of methodology: L. Yamamoto, I. Maier, J.-I. Yang, L.L. Presley, C. Chang, J. Borneman, R.H. Schiestl

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L. Yamamoto, I. Maier, A.T. Dang, D. Berry, J. Liu, R. Reliene, A.M. Westbrook, J. Borneman

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Yamamoto, I. Maier, A.T. Dang, D. Berry, J. Liu, P.M. Ruegger, J.-I. Yang, P.A. Soto, R. Reliene, A.M. Westbrook, A. Loy, J. Braun, J. Borneman, R.H. Schiestl

Writing, review, and/or revision of the manuscript: L. Yamamoto, I. Maier, A.T. Dang, D. Berry, J. Liu, P.M. Ruegger, L.L. Presley, R. Reliene, A. Loy, J. Braun, J. Borneman, R.H. Schiestl

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): I. Maier, P.M. Ruegger, J.-I. Yang, B. Wei, J. Braun, J. Borneman, R.H. Schiestl

Study supervision: I. Maier, J. Borneman, R.H. Schiestl

Acknowledgments

The authors thank N. Carls, D. Cowan and L. Parfenova for technical assistance.

Grant Support

This work was supported by NIH grants ES09519 (R.H. Schiestl), CA133928 (R. Reliene), AI078885 (J. Borneman), DK46763 (J. Braun), P30 CA016042 (Jonsson Comprehensive Cancer Center), the Crohn's and Colitis Foundation of America 3153 (J. Braun), the Broad Center of Regenerative Medicine and Stem Cell Research (J. Braun), the Austrian Federal Ministry of Science and Research (GEN-AU III InflammoBiota; D. Berry, A. Loy), NASA grant NNX11AB44G (R.H. Schietl, J. Borneman), the Ruzic Research Foundation (J. Braun), and dissertation fellowships from UC TSR&TP and UCLA Graduate Division (L. Yamamoto). J. Braun, J. Borneman, and R.H. Schiestl are filing a patent application related to this study.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 10, 2013; revised April 4, 2013; accepted April 27, 2013; published online July 16, 2013.

References

1. Meyn MS. Ataxia-telangiectasia, cancer and the pathobiology of the ATM gene. Clin Genet 1999;55:289–304.
2. Peterson RD, Funkhouser JD, Tuck-Muller CM, Gatti RA. Cancer susceptibility in ataxia-telangiectasia. Leukemia 1992;6 Suppl 1: 8–13.
3. Sandoval C, Swift M. Treatment of lymphoid malignancies in patients with ataxia-telangiectasia. Med Pediatr Oncol 1998;31:491–7.
4. Uziel T, Savitsky K, Platzer M, Ziv Y, Helbitz T, Nehls M, et al. Genomic organization of the ATM gene. Genomics 1996;33:317–20.
5. Rotman G, Shiloh Y. ATM: from gene to function. Hum Mol Genet 1998;7:1555–63.
6. Barlow C, Hirotsune S, Paylor R, Liyanage M, Eckhaus M, Collins F, et al. Atm-deficient mice: a paradigm of ataxia telangiectasia. Cell 1996;86:159–71.
7. Borghesani PR, Alt FW, Bottaro A, Davidson L, Aksoy S, Rathbun GA, et al. Abnormal development of Purkinje cells and lymphocytes in Atm mutant mice. Proc Natl Acad Sci USA 2000;97:3336–41.
8. Reliene R, Schiestl RH. Antioxidant N-acetyl cysteine reduces incidence and multiplicity of lymphoma in Atm deficient mice. DNA Repair (Amst) 2006;5:852–9.
9. Reliene R, Schiestl RH. Differences in animal housing facilities and diet may affect study outcomes-a plea for inclusion of such information in publications. DNA Repair (Amst) 2006;5:651–3.
10. Rao GN, Crockett PW. Effect of diet and housing on growth, body weight, survival and tumor incidences of B6C3F1 mice in chronic studies. Toxicol Pathol 2003;31:243–50.
11. Rescigno M. The pathogenic role of intestinal flora in IBD and colon cancer. Curr Drug Targets 2008;9:395–403.
12. Grivennikov SI, Wang K, Mucida D, Stewart CA, Schnabl B, et al. Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. Nature 2012;491:254–8.
13. Ward JM, Anver MR, Haines DC, Benveniste RE. Chronic active hepatitis in mice caused by *Helicobacter hepaticus*. Am J Pathol 1994;145:959–68.
14. Wu S, Rhee KJ, Albesiano E, Rabizadeh S, Wu X, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. Nat Med 2009;15:1016–22.
15. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 2012;338:120–3.
16. Pool-Zobel BL, Neudecker C, Domizlaff I, Ji S, Schillinger U, Rumney C, et al. Lactobacillus- and bifidobacterium-mediated antigenotoxicity in the colon of rats. Nutr Cancer 1996;26:365–80.
17. Fujiwara D, Wei B, Presley LL, Brewer S, McPherson M, Lewinski MA, et al. Systemic control of plasmacytoid dendritic cells by CD8+ T cells and commensal microbiota. J Immunol 2008;180:5843–52.
18. Rakoff-Nahoum S, Paglino J, Eslami-Varzaneh F, Edberg S, Medzhitov R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. Cell 2004;118:229–41.
19. Bishop AJ, Barlow C, Wynshaw-Boris AJ, Schiestl RH. Atm deficiency causes an increased frequency of intrachromosomal homologous recombination in mice. Cancer Res 2000;60:395–9.
20. Westbrook AM, Wei B, Braun J, Schiestl RH. Intestinal mucosal inflammation leads to systemic genotoxicity in mice. Cancer Res 2009;69:4827–34.
21. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Meth 2010;7:335–6.
22. Presley LL, Wei B, Braun J, Borneman J. Bacteria associated with immunoregulatory cells in mice. Appl Environ Microbiol 2010;76: 936–1.
23. Guo Z, Kozlov S, Lavin MF, Person MD, Paull TT. ATM activation by oxidative stress. Science 2010;330:517–21.
24. Wei B, Wingender G, Fujiwara D, Chen DY, McPherson M, Brewer S, et al. Commensal microbiota and CD8+ T cells shape the formation of invariant NKT cells. J Immunol 2010;184:1218–26.
25. Kullisaar T, Zilmer M, Mikelsaar M, Vihalemm T, Annuk H, Kairane C, et al. Two antioxidative lactobacilli strains as promising probiotics. Int J Food Microbiol 2002;72:215–24.
26. Federico A, Morgillo F, Tuccillo C, Ciardiello F, Loguercio C. Chronic inflammation and oxidative stress in human carcinogenesis. Int J Cancer 2007;121:2381–6.
27. Klaunig JE, Kamendulis LM, Hocevar BA. Oxidative stress and oxidative damage in carcinogenesis. Toxicol Pathol 2010;38:96–109.
28. Westbrook AM, Wei B, Hacke K, Xia M, Braun J, Schiestl RH. The role of tumour necrosis factor-α and tumour necrosis factor receptor signalling in inflammation-associated systemic genotoxicity. Mutagenesis 2012;27:77–86.
29. Aranda R, Sydora BC, McAllister PL, Binder SW, Yang HY, Targan SR, et al. Analysis of intestinal lymphocytes in mouse colitis mediated by transfer of CD4+, CD45RBhigh T cells to SCID recipients. J Immunol 1997;158:3464–73.
30. Chang C, Miller JF. *Campylobacter jejuni* colonization of mice with limited enteric flora. Infect Immun 2006;74:5261–71.
31. Wei B, Su TT, Dalwadi H, Stephan RP, Fujiwara D, Huang TT, et al. Resident enteric microbiota and CD8+ T cells shape the abundance of marginal zone B cells. Eur J Immunol 2008;38:3411–25.
32. Westbrook AM, Wei B, Braun J, Schiestl RH. Intestinal inflammation induces genotoxicity to extraintestinal tissues and cell types in mice. Int J Cancer 2011;129:1815–25.
33. van Baarlen P, Troost F, van der Meer C, Hooiveld G, Boekschoten M, Brummer RJ, et al. Microbes and Health Sackler Colloquium: human mucosal *in vivo* transcriptome responses to three lactobacilli indicate how probiotics may modulate human cellular pathways. Proc Natl Acad Sci USA 2011;108 Suppl 1:4562–9.
34. Jounai K, Ikado K, Sugimura T, Ano Y, Braun J, Fujiwara D. Spherical lactic acid bacteria activate plasmacytoid dendritic cells immunomodulatory function via TLR9-dependent crosstalk with myeloid dendritic cells. PLoS One 2012:7:e32588.
35. Vanderpool C, Yan F, Polk DB. Mechanisms of probiotic action: implications for therapeutic applications in inflammatory bowel diseases. Inflamm Bowel Dis 2008;14:1585–96.
36. Sgouras DN, Panayotopoulou EG, Martinez-Gonzalez B, Petraki K, Michopoulos S, Mentis A. *Lactobacillus johnsonii* La1 attenuates *Helicobacter pylori*-associated gastritis and reduces levels of proinflammatory chemokines in C57BL/6 mice. Clin Diagn Lab Immunol 2005;12:1378–86.
37. Isobe H, Nishiyama A, Takano T, Higuchi W, Nakagawa S, Taneike I, et al. Reduction of overall *Helicobacter pylori* colonization levels in the stomach of Mongolian gerbil *by Lactobacillus johnsonii* La1 (LC1) and

its *in vitro* activities against *H. pylori* motility and adherence. Biosci Biotechnol Biochem 2012;76:850–2.
38. La Ragione RM, Narbad A, Gasson MJ, Woodward MJ. *In vivo* characterization of *Lactobacillus johnsonii* FI9785 for use as a defined competitive exclusion agent against bacterial pathogens in poultry. Lett Appl Microbiol 2004;38:197–205.
39. Humen MA, De Antoni GL, Benyacoub J, Costas ME, Cardozo MI, Kozubsky L, et al. *Lactobacillus johnsonii* La1 antagonizes *Giardia intestinalis in vivo*. Infect Immun 2005;73:1265–9.
40. Valladares R, Sankar D, Li N, Williams E, Lai KK, Abdelgeliel AS, et al. *Lactobacillus johnsonii* N6.2 mitigates the development of type 1 diabetes in BB-DP rats. PLoS One 2010;5:e10507.
41. Joo HM, Hyun YJ, Myoung KS, Ahn YT, Lee JH, Huh CS, et al. *Lactobacillus johnsonii* HY7042 ameliorates *Gardnerella vaginalis*-induced vaginosis by killing *Gardnerella vaginalis* and inhibiting NF-κB activation. Int Immunopharmacol 2011;11:1758–65.
42. Parsonnet J. Bacterial infection as a cause of cancer. Environ Health Perspect 1995;103 Suppl 8:263–8.
43. Kaser A, Lee AH, Franke A, Glickman JN, Zeissig S, Tilg H, et al. XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. Cell 2008;134:743–56.
44. Todd DJ, Lee AH, Glimcher LH. The endoplasmic reticulum stress response in immunity and autoimmunity. Nat Rev Immunol 2008;8:663–74.
45. Pridmore RD, Berger B, Desiere F, Vilanova D, Barretto C, Pittet AC, et al. The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. Proc Natl Acad Sci USA 2004;101:2512–7.
46. Roos S, Jonsson H. A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components. Microbiology 2002;148:433–42.
47. Bensing BA, Sullam PM. An accessory sec locus of *Streptococcus gordonii* is required for export of the surface protein GspB and for normal levels of binding to human platelets. Mol Microbiol 2002;44:1081–94.
48. Stephenson AE, Wu H, Novak J, Tomana M, Mintz K, Fives-Taylor P. The Fap1 fimbrial adhesin is a glycoprotein: antibodies specific for the glycan moiety block the adhesion of *Streptococcus parasanguis* in an *in vitro* tooth model. Mol Microbiol 2002;43:147–57.
49. Allison GE, Fremaux C, Klaenhammer TR. Expansion of bacteriocin activity and host range upon complementation of two peptides encoded within the lactacin F operon. J Bacteriol 1994;176:2235–41.
50. Kingma SD, Li N, Sun F, Valladares RB, Neu J, Lorca GL. *Lactobacillus johnsonii* N6.2 stimulates the innate immune response through Toll-like receptor 9 in Caco-2 cells and increases intestinal crypt Paneth cell number in biobreeding diabetes-prone rats. J Nutr 2011;141:1023–8.

Chapter 3

Gut microflora and probiotic treatment affect radiation-induced genotoxicity

Chapter 3 is adapted from "Maier I, Liu J, Ruegger PM, Borneman J, and Schiestl RH. Intestinal bacterial phylotype indicators convey radiation-induced genotoxicity in microbiota-restricted and conventional mice. (in preparation)"

Abstract

The intestinal microbiota is thought to modulate carcinogenesis from endogenous sources as well as exogenous sources, such as ionizing radiation. We previously reported differences in radiation-induced systemic genotoxicity in mice raised under distinct microflora conditions. A probiotic *Lactobacillus johnsonii* strain, 456, was also shown to reduce genotoxicity in a DNA repair deficient mouse model. Here we test the oncoprotective effects of this strain on wild-type mice and under conditions of radiation exposure. Probiotic treatment reduced basal genotoxicity in conventionally-raised wild-type mice but exacerbated radiation-induced genotoxicity. The abundance of specific mucosal taxa also differed between mice exposed to different microflora and irradiation conditions, as well as probiotic treatment. Finally, irradiation inflicted higher DNA damage on mice raised in restricted microflora conditions compared to conventional housing, suggesting that protection or predisposition to genotoxicity by the gut microbiota and probiotics may be context-dependent.

Introduction

The gut microbiota has been found to play a role in determining an individual's intrinsic susceptibility to cancer. Besides studies of known oncogenic pathogens such as *Helicobacter pylori* (20) in the development of gastric cancer, it has been reported that gut microbial disturbances during colitis can initiate colorectal cancer (21). Previously, we also observed microflora-associated differences in systemic genotoxicity and lymphoma incidence using a mouse model with a genetic DNA repair deficiency (22).

Few studies have examined whether the gut microbiota can also modulate the effects of exogenous carcinogenic factors, such as exposure to ionizing radiation. Ionizing radiation inflicts both a physiological insult in the form of direct damage to cells and tissues as well as oncogenic

risk to surviving cells that have sustained DNA alterations. Known clinical effects of acute radiation exposure, such as diarrhea and mucositis, have been shown to be affected by the gut microbiota (23, 24), and a previous study conducted in our lab found differences in radiation-induced genotoxicity in mice raised under different microflora (25).

In this study, we evaluate the effect of a probiotic bacterium, *Lactobacillus johnsonii* 456, on radiation-induced genotoxicity. We previously observed that 456 could reduce systemic inflammation and genotoxicity in our DNA repair deficient mouse model (22), but it is unclear whether these oncoprotective effects extend to wild-type mice and to exogenous sources of genotoxicity, such as radiation.

Methods

Animal housing and husbandry

Wild-type C57/BL6J from the Jackson Laboratory (Bar Harbor, ME, USA) were housed under two types of specific pathogen free conditions, restricted microbiota (RM) and conventional microbiota (CM) (22), in the Division of Laboratory and Animal Medicine vivariums at the University of California, Los Angeles (UCLA, Los Angeles, CA). RM mice were created from a colony that was originally re-derived by Caesarian section and inoculated with only a few bacterial species as previously described by Fujiwara et al (26). CM mice were treated with antibiotics (27) followed by orogastric gavage of CM feces (22). In accordance with the Animal Research Committee at UCLA, either sterile (for RM), or non-sterile (for CM) food, water and bedding were used. Both mouse colonies were maintained under strictly aseptic and sterile conditions at the start of the experiments described below.

L. johnsonii culturing and suspension preparation

L. johnsonii strain 456 was isolated from RM wild-type mouse feces using Lactobacillus Selection Agar (BD Biosciences, San Jose, CA). For oral administration experiments, 456 was inoculated from cryogenic stocks onto BBL LBS Agar and incubated anaerobically for 2 days at 37°C. Bacterial growth was collected and washed three times by centrifugation and resuspension in phosphate-buffered saline (PBS). A final resuspension in 1 mL PBS was adjusted to a density of 10^9 colony forming units (CFU)/50 µL as determined by optical density (OD) readings of serial dilutions at 600 nm (equals OD=0.8 for our setting).

***L. johnsonii* inoculation and experimental time points of radiation experiments**

CM and RM mice for inoculation and radiation exposure studies were born in the respective colonies and were not treated with antibiotics. CM and RM mice were shipped from UCLA to Brookhaven National Laboratory (BNL, Upton, NY) and exposed to high-LET protons at 2.0 GeV/n provided by the NASA Space Radiation Laboratory. The total dose for each whole body irradiation experiment was 100-150 cGy delivered homogenously over 4-10 minutes with a dose rate of 12-15cGy/min. Six animals at a time were restrained in conical plastic holders for each exposure.

5-6 male CM and RM mice older than 6 months were orogastrically gavaged with suspensions of strain 456 (see above) at 1 day before and 2 days after irradiation. In addition, the drinking water for the same groups of animals was supplemented with an additional 10^9 CFU of the strain.

Female wild-type RM mice were at the age of 4-6 months when they were orally gavaged with 456 or PBS. Male and female mice for other radiation experiments were at the age of 3-4

months. Untreated mice were also transported from UCLA to BNL (males and females) and compared with mice maintained in the animal facilities at UCLA.

RM mice got autoclaved, acidified but sterile drinking water and autoclaved high-fat diet (crude fat 11%, PMI Nutrition International; Brentwood, MO), whereas CM mice got non-autoclaved water and non-autoclaved food. After irradiation, both CM and RM mice got acidified regular but sterile drinking water, and sterile food. All 456-inoculated mice received acidified drinking water and sterile food (9.6% fat, Prolab RMH 2000 5P06; Labdiet, St. Louis, MO).

Peripheral blood was drawn from the facial vein at more than 4-6 weeks after irradiation to examine lymphocyte genotoxicity via immunofluorescence staining of phosphorylated γ-H2AX as previous described (25). After another 3-4 months, mice were euthanized using 3% isoflurane and bacterial DNA was extracted from small intestine mucosa cells.

Intestinal bacterial analysis

Illumina-based high-throughput sequencing of bacterial 16S rRNA genes from intestinal mucosa samples (28) was performed with modifications as previously described (29). DNA was extracted from these samples using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA), and a 30-second bead-beating step using a Mini-Beadbeater-16 (BioSpec Products, Bartlesville, OK).

Sequence processing

Illumina sequencing data were processed with various scripts from Quantitative Insights Into Microbial Ecology (QIIME), version 1.7.0 (30). QIIME's split_libraries_fastq.py script was applied to de-multiplex and remove low-quality sequences using default quality parameters,

except that the minimum number of consecutive high-quality base calls to include a read as a fraction of input length was set to 0.99 (default = 0.75). All filtered reads had a length of 101 bp. The total initial number of sequencing reads was 232,332,834. Numbers of sequences removed by quality parameters: barcode not in mapping file (122,679,717), reads too short after quality truncation (7,198,500), and one or more Ns (104,671), leaving a total of 102,349,946 filtered reads (775,378 +/- 198,479 (mean +/- s.d.) per sample).

De novo operational taxonomic units (OTUs) were picked at 97% sequence similarity with UCLUST (31), resulting in 297,163 OTUs. Smaller OTUs (defined as those containing 1,000 reads or less) were removed, leaving 1,767 OTUs for downstream processing. Putative taxonomic classification of OTUs was performed with the assign_taxonomy.py script using the BLAST option against the May 2013 version of the Greengenes database (32), pre-clustered at 97% sequence similarity.

Results

A probiotic *L. johnsonii* strain reduces basal genotoxicity in wild-type mice

L. johnsonii strain 456 was previously found to reduce systemic genotoxicity in DNA repair-deficient mice (22), so we first established its effectiveness in wild-type mice raised either in a conventional microflora (CM) or a restricted microflora (RM) that was described in previous studies (22, 25, 26). We inoculated 456 into CM and RM mice and measured the phosphorylation of histone protein γ-H2AX, a marker of DNA double-stranded breaks, in whole blood lymphocytes 12 weeks after inoculation. 456 inoculation into RM mice did not clearly reduce the frequency of phosphorylated γ-H2AX foci relative to untreated mice (Figure 3-1, “Inhouse” groups). However, in CM mice, which show relatively higher inherent lymphocyte

genotoxicity, treatment with 456 reduced γ-H2AX phosphorylation to a level similar to that of RM mice (Figure 3-1).

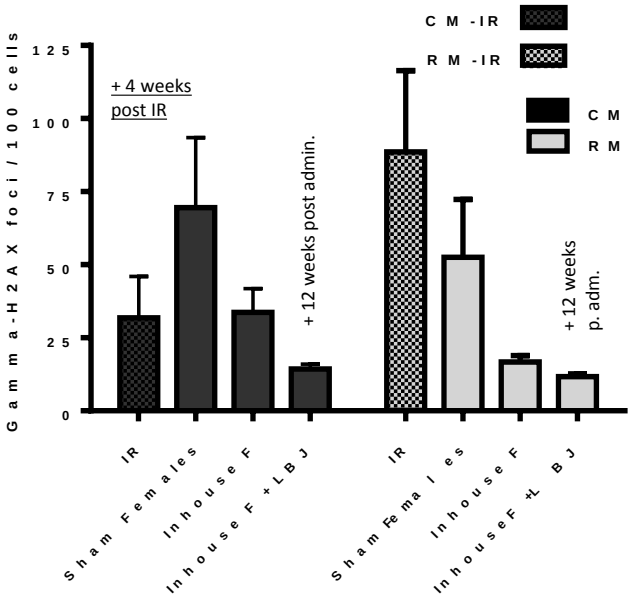


Figure 3-1. Strain 456, microbiota, and proton irradiation alter systemic genotoxicity in wild-type mice. Female CM (n=5) and RM (n=5) mice were proton-irradiated and γ-H2AX foci were assessed at four weeks post radiation and compared with unirradiated sham mice (n=3 for CM, n=4 for RM). γ-H2AX foci were also assessed at twelve weeks after administration of 456 to in-house female unirradiated mice (at age of 4-6 months; n=3 for CM and n=4 for RM) for comparison with untreated mice in the same facility (n=4 for CM and RM). Error bars indicate SEM. IR=Irradiation; LBJ=456-inoculated. Figure credited to Irene Maier.

RM mice sustain higher radiation-induced DNA damage than CM mice

We next evaluated the modulatory effects of 456 and microflora conditions on radiation-induced genotoxicity. We orogastrically inoculated 456 into CM and RM mice 1 day before and 2 days after exposure to 1 – 1.5 Gy high-energy protons and measured systemic lymphocyte genotoxicity four weeks after exposure. Groups of uninoculated RM and CM mice were also irradiated to determine the effect of microflora composition alone. In these latter groups, we observed a significantly higher frequency of phosphorylated γ-H2AX foci in the circulating lymphocytes of irradiated RM mice compared with irradiated CM mice ($p<0.0001$, Figure 3-2). Lymphocyte genotoxicity was not significantly different between irradiated mice and unirradiated mice under CM conditions, however, under RM conditions, irradiation resulted in a significant increase in γ-H2AX phosphorylation (Figure 3-2).

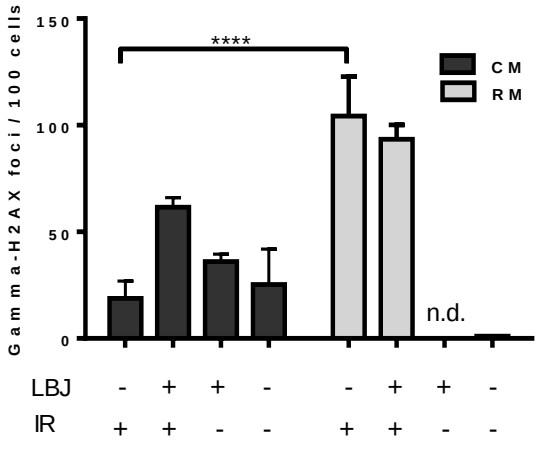


Figure 3-2. Double-stranded DNA breaks in blood lymphocytes of 456-inoculated mice and irradiated mice. Male CM and RM mice (at age >6 months; n=5 in each group) were gavaged with 456 before, and at two days post exposure to a sublethal dose of high-LET and -energy protons. n=5 for irradiated mice, n=3 for 456-inoculated mice, and n= 4 for untreated mice. IR=Irradiation, LBJ= 456-inoculated, n.d.=not determined, **** indicates $p<0.0001$. Adapted from figure credited to Irene Maier.

We used 16S rRNA sequencing to compare the fecal bacterial compositions of mice under different irradiation, 456 inoculation, and microflora conditions and found several differentially abundant taxa (Table 1). For example, an unclassified Gram-negative species cTPY-13 and *L. johnsonii* were more abundant in RM mice while an unclassified *Ureaplasma* species was more abundant in CM mice. These taxa generally changed in abundance after whole body irradiation or 456-inoculation, however it is unclear whether the differential abundances are due to irradiation, 456-inoculation, microflora, or an interaction of these factors.

Table 1. Small intestinal bacteria differentially abundant between irradiation, microflora, and inoculation conditions (% of reads)

Nearest Cultured Relative (accession) (% identity)	CM-IR	CM-L-IR	CM-L	CM	RM-IR	RM-L-IR	RM-L	RM	ANOVA p-value
Gram-negative bacterium cTPY-13 (AY239461) (94%)	25.2	1.8	1.0	7.9	17.2	32.2	18.7	49.6	0.006
Porphyromonadaceae bacterium S9 HS-8 (KF007168) (93%)	5.7	4.1	2.3	2.7	4.1	1.0	0.5	0.1	0.017
Gram-negative bacterium cTPY-13 (AY239461) (96%)	4.9	2.0	0.7	2.1	12.0	3.5	0.8	3.7	0.028
***Ureaplasma* sp. (JN713485) (96%)**	1.1	8.7	35.1	3.2	0.0	0.0	0.0	0.0	0.008
***Lactobacillus johnsonii* (CP006811) (100%)**	0.1	0.0	0.0	0.1	1.7	2.9	3.1	1.9	0.470

456 increases radiation-induced genotoxicity in CM mice but has no effect in RM mice

In RM mice, 456 inoculation did not significantly alleviate the increase in γ-H2AX phosphorylation 4 weeks after irradiation (Figure 3-2). Under CM conditions, mice receiving both 456 inoculation and irradiation unexpectedly sustained significantly higher lymphocyte γ-H2AX phosphorylation than mice receiving irradiation alone (Figure 3-2), suggesting that 456 may promote genotoxicity in these mice under conditions of ionizing radiation exposure.

Discussion

The gut microbiota is known to mediate as well as modulate some of the biological effects of ionizing radiation. Shifts in the gut microbiota in response to radiation have previously been described in human radiotherapy patients (33) as well as in mice (34). Moreover, the composition of the gut microbiota prior to irradiation may predispose certain radiotherapy patients to developing diarrhea (24) and thus has been investigated in this capacity as a prognostic (35). With regard to the genotoxic effects of ionizing radiation, a previous study in our lab demonstrated different levels of genotoxicity in irradiated mice raised in different microflora conditions (25). Consistent with this study, we find an increased susceptibility to proton radiation-induced genotoxicity in wild-type mice raised in RM compared with CM. That RM can promote radiation-induced genotoxicity stands in contrast to the previously observed association between RM and reduced basal genotoxicity (22). It is possible that protection against or promotion of genotoxicity by the microbial community represented by RM may be context-dependent.

This context dependency may also extend to probiotic manipulations of the gut microbiota. We find that while 456 can reduce systemic genotoxicity in CM mice, it may also exacerbate systemic DNA damage from proton radiation, a common type of radiation employed

in radiotherapy. Further investigation is needed to determine whether 456, and possibly *L. johnsonii* in general, may pose a cancer risk under conditions of irradiation exposure. The manipulation of the gut microbiota has been considered as a potential adjuvant in cancer treatment, as gut microbes and probiotic administration have already been demonstrated to influence the effectiveness of cancer immunotherapy (36, 37). Evidence also exists that they may reduce side effects of radiation exposure in mice and radiotherapy patients (38). Our findings in 456 suggest that caution may be warranted in the selection of probiotic adjuvants.

Chapter 4

Genome sequencing and comparative analysis of *Lactobacillus johnsonii* 456

Abstract

The gut commensal bacterium *Lactobacillus johnsonii* has been studied for its probiotic properties. Recent investigations have shown that host and bacterial genetic factors may influence *L. johnsonii* colonization, and, consequently, its effectiveness as a probiotic. We thus performed a genome comparison of 9 sequenced *L. johnsonii* strains from various hosts. Comparison of core genomic regions revealed that isolates from the same or similar hosts shared the highest similarity, with some exceptions. Non-core regions among the isolates contain many genes involved in propagation of, or defense against, mobile genetic elements, and the presence of lactocepin proteases was specific to the isolates from specific hosts, suggesting that phage selective pressure and metabolic/immunomodulatory capabilities may contribute to differences in the successful colonization of different *L. johnsonii* isolates.

Introduction

Lactobacillus johnsonii is a Gram-positive gut commensal bacterium commonly found in many vertebrate hosts, including humans. It has been studied for its potential probiotic properties, including the inhibition of intestinal pathogens and their associated diseases (15–17, 39) as well as anti-inflammatory or immunomodulatory interaction with host tissues (19, 40).

We previously isolated *L. johnsonii* strain 456 from a DNA repair-deficient mouse model that showed lower systemic genotoxicity and cancer risk when raised in restricted in a restricted microflora (22). Long-term inoculation of strain 456 into this mouse model under conventional housing conditions reduced systemic genotoxicity and inflammation, suggesting a potential application for this strain as an anti-inflammatory and oncoprotective probiotic in humans.

The effectiveness of *L. johnsonii* probiotics may depend on their compatibility with the recipient host, as suggested by recent work on the genetic basis of *L. johnsonii*-host mutualism. For instance, analysis of conserved gene markers and isolate genomes have revealed genetic distinctions among the *L. johnsonii* strains that inhabit humans, pigs, poultry, and mice (10, 41) as well as the influence of host genetics on *L. johnsonii* abundances in the gut (11). Several bacterial genes have also been identified that may mediate host attachment and strain-specific differences in the gut persistence of *L. johnsonii* isolates (42–44). To better understand the compatibility of strain 456 in the context of these genetic factors, we sequenced its genome and performed a genome-wide and gene-specific comparison with 8 other sequenced *L. johnsonii* isolates.

Methods

Sequencing and genome assembly of *L. johnsonii* strain 456

L. johnsonii 456 was cultured anaerobically in Lactobacillus Selection Broth (BD, Franklin Lakes, New Jersey, USA) at 37 °C for 2 days. Paired-end sequencing was performed on an Illumina HiSeq. Reads were assembled using Mira 4.0.2 (45) and manually finished in Gap5 (46). Contigs longer than 500 bases were scaffolded against the genome of *L. johnsonii* strain NCC533 using scaffold_builder (47). Open reading frames (ORFs) were annotated based on the genome of NCC 533 using the Rapid Annotation Transfer Tool (48). ORFs were also predicted using Genemark.hmm (49) and Glimmer 3.02 (50) and manually revised using Artemis (51).

Comparative genome analysis of *L. johnsonii* strains

I compared the genome of strain 456 with 8 other *L. johnsonii* genomes available from NCBI on April 10, 2016. I first calculated the core genomic regions shared by all 9 *L.johnsonii*

strains, as described in Tomida *et al.* (52). Briefly, Nucmer was used to identify homologous regions between the genome of NCC 533 and each of the other 8 genomes. The set of core genomic regions was determined to be the regions homologous to NCC 533 that were present in the genomes of the 8 other strains. Single nucleotide polymorphisms (SNPs) were identified within core regions using Nucmer and were used to construct a phylogenetic tree in MEGA 5 using the Neighbor-Joining method on p-distances (53). Bootstrapping was performed using 500 replicates.

Non-core genomic regions among the *L. johnsonii* genomes were also identified as described in Tomida *et al.* (52). Briefly, a pan-genome across all 9 *L. johnsonii* strains was constructed by first using Nucmer to comparing the NCC 533 genome with one of the 8 other genomes. Regions with no NCC 533 homologs were concatenated to the NCC 533 genome sequence. This concatenated sequence was then iteratively compared using the same method to each of the remaining genomes to construct the pan-genome. Finally, the pan-genome was compared to each of the 9 genomes individually to identify non-core regions that were absent in at least one of the genomes.

Lactocepins were identified in the genomes by first creating a database of all proteins in the NCBI RefSeq protein database containing the keyword "lactocepin". tblastn was used to identify homologous regions within the 9 *L. johnsonii* genomes. The translated amino acid sequences of genes containing these regions were aligned using MUSCLE (54).

Results

Sequencing and assembly of the *L. johnsonii* 456 genome

Assembly of 211,866 reads sequenced from *L. johnsonii* 456 genomic DNA resulted in a 1.58 Mb draft genome. The assembly consisted of 1,007 scaffolds and contigs, with an estimated coverage of 20× (N50: 2,421 nt). Reference-based and *de novo* gene prediction identified 1,818 genes and 28 tRNAs.

Analysis of core regions in 11 *L. johnsonii* genomes

We determined the genetic relatedness between 456 and 8 other sequenced *L. johnsonii* strains within the larger diversity of *L. johnsonii* characterized by Buhnik-Rosenblau et al (10). In this study, 46 *L. johnsonii* isolates from diverse hosts were genotyped at three conserved loci. We incorporated the corresponding homologous regions from each of the 9 *L. johnsonii* genomes into this analysis (Figure 4-1A) and found, consistent with the previous study, that the 9 sequenced isolates clustered with the other genotyped isolates based on host of origin. However, we noted some exceptions, such as the human isolates ATCC 33200 and NCC 1741, which respectively grouped among isolates from poultry and isolates from bovine and porcine hosts. 456, isolated from mice, also grouped separately from the other mouse isolates. Two genomes, 770_LJOH and 987_LJOH, were classified as belonging to *L. johnsonii* but formed a separate clade that was the most distantly related to all other *L. johnsonii* strains.

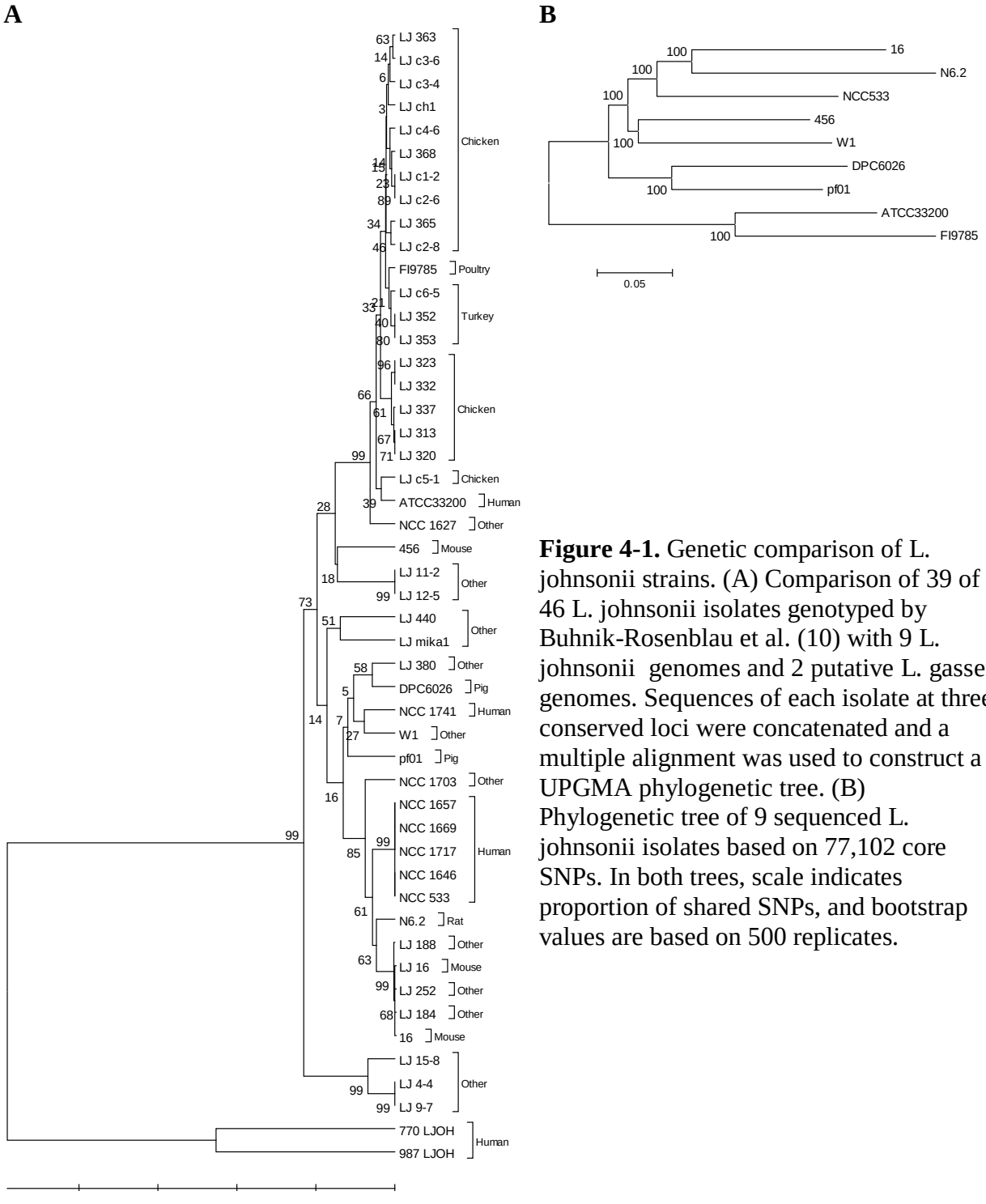


Figure 4-1. Genetic comparison of L. johnsonii strains. (A) Comparison of 39 of 46 L. johnsonii isolates genotyped by Buhnik-Rosenblau et al. (10) with 9 L. johnsonii genomes and 2 putative L. gasseri genomes. Sequences of each isolate at three conserved loci were concatenated and a multiple alignment was used to construct a UPGMA phylogenetic tree. (B) Phylogenetic tree of 9 sequenced L. johnsonii isolates based on 77,102 core SNPs. In both trees, scale indicates proportion of shared SNPs, and bootstrap values are based on 500 replicates.

We confirmed the phylogenetic relationships among *L. johnsonii* isolates at the genome level. Based on nucleotide similarity at 159,108 sites within the core genomic regions of the 9 strains (Figure 4-1B), DPC6026, a pig isolate, was most similar to another pig isolate, pf01, while a mouse isolate, strain 16, was most similar to a rat isolate, strain N6.2, consistent with the above analysis of the conserved loci. However, the human isolate ATCC 33200 was most similar to the poultry isolate FI9785 while strain 456 clustered separately from the other rodent isolates. As before, the genomes of 770_LJOH and 987_LJOH formed a clade distantly related to the other *L. johnsonii* strains. However, the genomes of four *Lactobacillus gasseri* strains also clustered within this clade when we included them in the analysis (Figure 4-2), warranting a possible reclassification of 770_LJOH and 987_LJOH as *L. gasseri* strains.

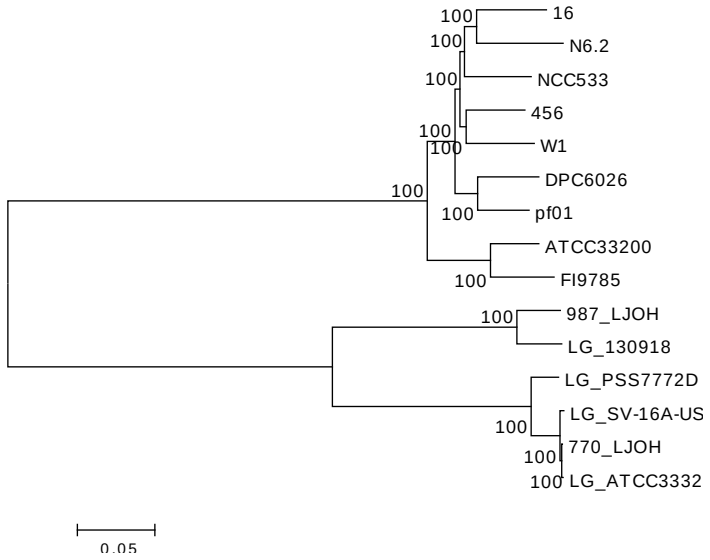


Figure 4-2. 770_LJOH and 987_LJOH may be strains of L. gasseri. Phylogenetic tree of 9 sequenced L. johnsonii isolates, 2 suspected L. gasseri isolates, and 4 L. gasseri isolates (indicated by "LG") based on 152,754 core SNPs. Scale indicates proportion of shared SNPs, and bootstrap values are based on 500 replicates.

Analysis of non-core regions in 11 *L. johnsonii* genomes

We next determined whether the *L. johnsonii* genomes may encode specific adaptations to conditions within different hosts. We determined the non-core genomic regions among the

sequenced isolates by comparing each genome to a pan-genome constructed from the unique regions across all 9 *L. johnsonii* genomes (Figure 4-3). The presence and absence of non-core regions were similar among isolates from similar hosts, however, several genome regions shared among all or most strains were not found in 456, setting it apart from the other isolates. Most of these missing regions corresponded to homologous regions in the NCC 533 genome that mapped to assembly gaps within 456, suggesting that they may be unsequenced areas of the 456 genome rather than true relative deletions in 456.

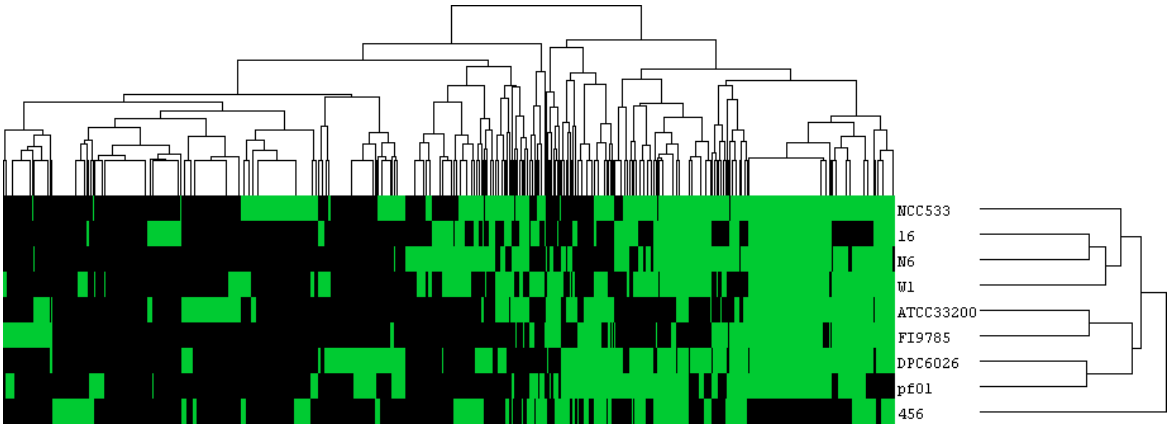


Figure 4-3. Non-core genomic regions of 9 L. johnsonii strains. Map indicates the presence (green) or absence (black) of 746 genomic regions >500 nt in each of the strains. Dendrograms are based on average linkage hierarchical clustering of Euclidean distances.

Each isolate genome contained between 1 – 30 non-core regions (cumulative length of 57,721 – 172,746 bp) that were not detected in the other genomes. These strain-specific non-core regions accounted for an average of 5% of the length of each sequenced genome. This is consistent with a previous study based on hybridization experiments (55). While many of the genes in these regions have unknown function, phage and mobile element proteins as well as restriction modification enzymes frequently occurred in the unique regions of the different strains (Table 4-1). 456 notably contained various cell-wall associated proteins, including 5

ORFs annotated as mucus-binding proteins, which may impart specific host attachment properties.

Table 4-1. Mobile element, phage defense, and mucus-binding proteins in the non-core regions specific to each strain.
Shown are the number of homologs of each category detected for each strain.

Protein	16	456	ATCC33200	DPC6026	FI9785	pf01	Total strains with function
cI repressor	0	0	1	0	0	0	1
conjugal transfer protein TraG	0	1	0	0	0	0	1
head protein	0	0	0	0	2	0	1
head-tail adaptor protein	1	0	0	0	0	0	1
holin	0	0	1	0	0	0	1
integrase	0	2	3	0	6	6	4
lysin	0	0	0	0	2	1	2
phage head morphogenesis protein	0	0	0	0	1	0	1
phage head protein	0	0	0	0	1	0	1
phage protein	0	0	0	0	0	1	1
phage related protein	0	0	0	0	1	0	1
phage replication protein	0	0	1	0	0	0	1
phage tail protein	1	0	0	0	1	0	2
phage tail tape measure protein	0	0	0	0	1	0	1
portal protein	1	0	0	0	0	0	1
terminase	1	0	0	0	1	0	2
toxin Bro	1	0	0	0	0	0	1
toxin YoeB	0	0	0	0	1	0	1
toxin-antitoxin system antitoxin subunit	0	0	0	0	1	0	1
transposase	2	0	9	3	15	11	5
abortive infection protein	0	0	1	0	1	0	2
abortive infection protein AbiGII	0	0	0	0	0	1	1
endonuclease	0	0	0	0	0	5	1
restriction endonuclease	0	2	0	0	1	1	3
restriction endonuclease subunit R	0	0	0	0	1	0	1
restriction endonuclease subunit S	0	0	1	0	0	0	1
type I restriction endonuclease subunit S	0	1	0	0	0	0	1
type I restriction-modification protein subunit M	0	0	0	0	0	1	1
type I restriction-modification system subunit M	0	1	0	0	0	0	1
type I restriction-modification system, subunit S	0	0	0	0	0	2	1
type II restriction endonuclease subunit M	0	0	1	0	1	0	2
type II restriction enzyme StsI	0	0	0	0	0	1	1
type III restriction endonuclease StyLTI	0	0	1	0	0	0	1
type III restriction endonuclease subunit M	0	0	0	0	1	0	1

type III restriction endonuclease subunit R	0	0	0	0	1	0	1
mucus binding protein	0	0	0	0	2	0	1
mucus binding protein precursor Mub	0	0	1	0	0	0	1
mucus-binding protein	0	6	0	0	0	0	1
Mucus-binding protein Mub	0	1	0	0	0	0	1

Putative lactocepins encoded by *L. johnsonii*

Lactic acid bacteria express a diversity of soluble and cell-envelope proteases (lactocepins) involved in the catabolic processing of proteins for uptake (56). Some lactocepins have been found to degrade inflammatory chemokines (57) and may thus explain the anti-inflammatory effects of certain lactobacilli, including *L. johnsonii* (19, 58, 59). To investigate whether different strains carry different lactocepins and thus have different anti-inflammatory potential, we analyzed the lactocepin sequences encoded in the 9 *L. johnsonii* genomes. We found 3 types of putative lactocepins (Table 4-2). The first type was found in strains from humans (NCC 533), rodents (16, N6.2, 456), and an unclear host (W1), and represents a distant homolog of lactocepins found in *Streptococcus gallolyticus* (28-37% identity). The second type was found only within strains NCC 533, 456, and 16, and represents a subtilisin-like S8 peptidase. Finally, the third type corresponds to a distant homolog of a lactocepin in *Lactobacillus acetotolerans*, and was found only in W1. No homologs were found in the genomes of pig or poultry isolates despite the completeness of the DPC 6026 and FI9785 genomes.

Table 4-2. *L. johnsonii* lactocepins

Protein ID	Isolate	Length (aa)	Homology	Similarity
T285_RS04970	N6.2	2,103	*Streptococcus gallolyticus* lactocepins	28 – 37% identity
456_LJ_RS04945	456	2,728		
LJ16_RS06075	16	2,760		
LJ_RS04945	NCC 533	4,734		
AYJ53_RS06045	W1	4,995		
456_LJ_RS09445	456	2,242	subtilisin-like S8 peptidase	23 – 63% identity
LJ_RS09445	NCC 533	2,209		
AYJ53_RS04080	W1	2,273		
AYJ53_RS02405	W1	682	*Lactobacillus acetotolerans* lactocepin	45% identity to peptidase domain, 26% identity to SLAP domain

Discussion

The genetic differences between *L. johnsonii* strains from different organisms are currently thought to reflect an adaptive co-evolution of this gut commensal bacterium with different hosts (10, 41). Consistent with these studies, we find that the core regions of 9 currently available *L. johnsonii* genomes showed highest nucleotide similarity between isolates from the same or similar hosts. However, we also observe that some isolates are genetically distant from others from the same host, based on comparison of conserved loci and core genomic regions. Some of these discrepancies may be explained by the possibly transient transmission of *L. johnsonii* between interacting hosts, such as from prey consumed by predators, which has been suggested to explain the close genetic similarity between isolates from mice and from a few of their natural predators (10). This explanation may also account for the close similarities between bovine, porcine, and poultry livestock isolates with the human-isolated strains ATCC 33200 and NCC 1741.

Another explanation is that strains from these host-specific genetic groups may be capable of colonizing other hosts. In our study, we observe that strain 456 is genetically distant

from the other *L. johnsonii* mouse isolates, despite having been isolated from mice raised in restricted flora. As mice housed under these conditions are maintained on sterilized food and bedding and are handled using stringent aseptic techniques, 456 is likely a stable gut commensal in these mice rather than a transient environmental contaminant. Limited studies of *L. johnsonii* colonization in non-native hosts have also demonstrated that the human isolate NCC 533 can persist for longer durations in the mouse gut than another human isolate, ATCC 33200 (43), which we have shown is more genetically similar to poultry isolates. However, further studies are needed to more clearly define the host range of *L. johnsonii* strains.

The differential gut persistence of *L. johnsonii* strains may be mediated by several genetic adaptations. Genes involved in host attachment (42, 44) have already been identified as possible factors for stable colonization. Here, our analysis of strain-specific non-core regions identified genes involved in the propagation of or defense against phages and mobile elements, raising the possibility that the phages native to different hosts may be a another selective pressure that affects the persistence of certain *L. johnsonii* isolates. We also found lactocepins within the genomes of human and rodent *L. johnsonii* isolates but not in two complete pig and poultry isolate genomes. Further investigation is needed to confirm the general absence of these proteases in poultry and pig isolates as well as the potential metabolic or immunological consequences. The presence of these lactocepin homologs in 456, along with the enrichment of genes encoding cell-surface, mucus-binding proteins within its strain-specific regions, may explain the persistence of this strain in the mouse gut, despite its lack of similarity to other mouse isolates.

Chapter 5

The diversity and host interactions of *Propionibacterium acnes* bacteriophages on human skin

The ISME Journal (2015), 1–16
© 2015 International Society for Microbial Ecology All rights reserved 1751-7362/15
www.nature.com/ismej

ORIGINAL ARTICLE

The diversity and host interactions of *Propionibacterium acnes* bacteriophages on human skin

Jared Liu[1], Riceley Yan[1], Qiao Zhong[1,2], Sam Ngo[1], Nathanael J Bangayan[1], Lin Nguyen[1], Timothy Lui[1], Minghsun Liu[3], Marie C Erfe[4], Noah Craft[4], Shuta Tomida[1,6] and Huiying Li[1,5]
[1]*Department of Molecular and Medical Pharmacology, Crump Institute for Molecular Imaging, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA;* [2]*Department of Laboratory Medicine, Suzhou Municipal Hospital, Suzhou Hospital Affiliated to Nanjing Medical University, Suzhou, China;* [3]*Department of Microbiology, Immunology and Molecular Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA;* [4]*Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Los Angeles, CA, USA and* [5]*UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA, USA*

The viral population, including bacteriophages, is an important component of the human microbiota, yet is poorly understood. We aim to determine whether bacteriophages modulate the composition of the bacterial populations, thus potentially playing a role in health or disease. We investigated the diversity and host interactions of the bacteriophages of *Propionibacterium acnes*, a major human skin commensal implicated in acne pathogenesis. By sequencing 48 *P. acnes* phages isolated from acne patients and healthy individuals and by analyzing the *P. acnes* phage populations in healthy skin metagenomes, we revealed that *P. acnes* phage populations in the skin microbial community are often dominated by one strain. We also found phage strains shared among both related and unrelated individuals, suggesting that a pool of common phages exists in the human population and that transmission of phages may occur between individuals. To better understand the bacterium–phage interactions in the skin microbiota, we determined the outcomes of 74 genetically defined *Propionibacterium* strains challenged by 15 sequenced phages. Depending on the *Propionibacterium* lineage, phage infection can result in lysis, pseudolysogeny, or resistance. In type II *P. acnes* strains, we found that encoding matching clustered regularly interspaced short palindromic repeat spacers is insufficient to confer phage resistance. Overall, our findings suggest that the prey–predator relationship between bacteria and phages may have a role in modulating the composition of the microbiota. Our study also suggests that the microbiome structure of an individual may be an important factor in the design of phage-based therapy.
The ISME Journal advance online publication, 7 April 2015; doi:10.1038/ismej.2015.47

Introduction

The human skin is inhabited by hundreds of microbial species, including bacteria, fungi and viruses. The homeostasis of this ecosystem is important to its function as a barrier against infection and colonization of pathogens on the skin surface. Bacteriophages are important components of the human microbiota. They are a reservoir of diversity-generating elements (Rohwer and Thurber, 2009) and regulate both the abundance and diversity of their bacterial hosts by predation (Suttle *et al.*, 1990; Waterbury and Valois, 1993; Rohwer, 2003; Rodriguez-Valera *et al.*, 2009). Although the skin bacterial community has been studied by several groups (Gao *et al.*, 2007; Costello *et al.*, 2009; Grice *et al.*, 2009; Kong *et al.*, 2012; The Human Microbiome Project Consortium, 2012; Blaser *et al.*, 2013; Fitz-Gibbon *et al.*, 2013; Nakatsuji *et al.*, 2013), relatively few studies have characterized the skin viral community (Foulongne *et al.*, 2012; Ma *et al.*, 2014; Wylie *et al.*, 2014). In particular, the composition and dynamics of bacteriophages and their interactions with bacterial hosts on the skin are not well understood.

The microbial community in the skin pilosebaceous unit is dominated by *Propionibacterium acnes*, which accounts for nearly 90% of the microbiota (Fitz-Gibbon *et al.*, 2013). Although *P. acnes* is a major skin commensal, it has been considered a pathogenic factor for acne vulgaris (Leyden, 2001; Bojar and Holland, 2004), one of the most common skin diseases affecting over 80% of adolescents

Correspondence: H Li, Department of Molecular and Medical Pharmacology, Crump Institute for Molecular Imaging, David Geffen School of Medicine, UCLA, 4339 CNSI, 570 Westwood Plaza, Building 114, Los Angeles, CA 90095-1770, USA.
E-mail: huiying@mednet.ucla.edu
[6]Current address: Department of Genome Biology, Kinki University, Osaka, Japan.
Received 9 October 2014; revised 12 February 2015; accepted 26 February 2015

and young adults (White, 1998; Bergler-Czop and Brzezińska-Wcisło, 2013). Our previous 16S ribosomal RNA metagenomic study demonstrated that *P. acnes* strain population structure in pilosebaceous units differs significantly between acne patients and healthy individuals (Fitz-Gibbon *et al*., 2013). Certain strains are highly associated with the disease (Lomholt and Kilian, 2010; McDowell *et al*., 2012; Fitz-Gibbon *et al*., 2013; Tomida *et al*., 2013), while some strains are enriched in healthy skin (Fitz-Gibbon *et al*., 2013).

In parallel, *P. acnes* phages are dominant bacteriophages in the pilosebaceous unit (Fitz-Gibbon *et al*., 2013). It has been known for over 50 years that *P. acnes* phages exist on the human skin (Brzin, 1964). They have the morphology of siphoviruses, consisting of a ~50 nm icosahedral head and a ~150 nm flexible tail (Farrar *et al*., 2007). Zierdt *et al*., (1968) isolated phage 174 from spontaneous plaques of a *P. acnes* isolate. Phage 174 was able to lyse nearly all *P. acnes* strains tested in the study. Subsequently, more *P. acnes* phages were isolated, which exhibited lytic as well as pseudolysogenic behavior (Lood and Collin, 2011). However, in the past decades, the study of *P. acnes* phages had been limited to the development of phage typing systems to distinguish different serotypes of *P. acnes* (Jong *et al*., 1975; Webster and Cummins, 1978). Recent sequencing of 14 *P. acnes* phages (Farrar *et al*., 2007; Lood and Collin, 2011; Marinelli *et al*., 2012) suggested that they have limited genetic diversity with over 85% nucleotide identity in the genome. All sequenced genomes are similar in size and structure with 45–47 genes encoded in two oppositely transcribed regions named the left arm and right arm (Farrar *et al*., 2007; Lood and Collin, 2011; Marinelli *et al*., 2012).

Much is to be learned about whether bacteriophages drive the diversity and dynamics of the skin bacterial community. The ratio between *P. acnes* phages and *P. acnes* was ~1:20 in pilosebaceous units, based on a pooled healthy skin sample that we analyzed previously (Fitz-Gibbon *et al*., 2013), but can vary in a large range among individuals and over time. *P. acnes* phages do not encode integrases in their genomes (Farrar *et al*., 2007; Lood and Collin, 2011), suggesting their inability to stably integrate into the host chromosome. They can kill the host bacteria through cell lysis or can enter a pseudolysogenic state in the host strain (Lood and Collin, 2011), in which the phage DNA persists in infected cells without lysing the host or integrating into its genome. Whether *P. acnes* phages modulate the relative abundances of different *P. acnes* strains by selectively killing specific strains of *P. acnes* and thus play a role in skin health and disease is unknown.

On the other hand, bacterial hosts can influence phage populations through antiviral mechanisms, such as the restriction modification mechanism and the bacterial adaptive immune system utilizing clustered regularly interspaced short palindromic repeat (CRISPR) sequence arrays (Horvath and Barrangou, 2010). In our effort to characterize the strain diversity of *P. acnes* in the pilosebaceous unit, we discovered that all sequenced type II *P. acnes* strains harbor Type I-E CRISPR and CRISPR-associated (Cas) proteins (Fitz-Gibbon *et al*., 2013; Tomida *et al*., 2013). Marinelli *et al*. (2012) suggested that the CRISPR mechanism explains the resistance of certain *P. acnes* strains to phage infection, yet noting that some of their observations were inconsistent with this theory.

To better understand how bacteriophages modulate the bacterial composition of the skin microbiota and their potential roles in skin health and disease, in this study, we determined the diversity of *P. acnes* phages and their interactions with bacterial hosts in the skin of acne patients and healthy individuals. We sequenced the genomes of 48 *P. acnes* phages isolated from 37 individuals and investigated whether certain phage strains dominate the skin microbiota. By analyzing the skin metagenome data from the Human Microbiome Project (HMP), we further verified our conclusions from analyzing sequenced phage isolates. We also challenged a panel of 74 genetically defined *Propionibacterium* strains against 15 of the sequenced phages to determine the outcome and mechanisms of their interactions.

Materials and methods

Phage isolation and DNA extraction

Skin follicle samples were previously collected from the nose of acne patients and individuals with healthy skin as reported in the study by Fitz-Gibbon *et al*., (2013). To best represent the diversity of populations and history of medical care, the subjects were recruited from private practice, managed care and public hospital settings, as well as outside of dermatology clinics in Southern California. Written informed consent was provided by all study subjects.

The follicle contents collected on the surface of the nose strip were mashed using a sterile loop (Fisherbrand, Pittsburgh, PA, USA), and plated onto a blood agar plate (Teknova Brucella Agar Plate with Hemin and Vitamin K, Teknova, Hollister, CA, USA). The sample plates were incubated at 37 °C for 5–7 days anaerobically using the AnaeroPack System (Mitsubishi Gas Chemical Company, Tokyo, Japan) (Fitz-Gibbon *et al*., 2013).

Phage plaques observed on the culture plates were isolated by puncturing the agar with a sterile pipette tip and resuspending each tip in 50 µl SM buffer (0.1 M sodium chloride, 8 mM magnesium sulfate heptahydrate, 1M Tris-HCl, pH 7.5, 2% gelatin and 1 mM calcium chloride). Each phage resuspension was spread onto A-media plates (12 g l^{-1} pancreatic digest of casein, 12 g l^{-1} yeast extract, 22.2 mM

D-glucose, 29.4 mM g l^{-1} potassium phosphate monobasic, 8 mM magnesium sulfate heptahydrate and 20 g l^{-1} agar) with top agar containing *P. acnes* strain ATCC6919. After incubation at 37 °C for 2 days, single plaques were selected and propagated using the same host strain, medium and incubation conditions. Suspensions of each phage isolate were prepared by eluting plates with 8 ml SM buffer at room temperature, filtering with a 0.22-μm PES filter (Millipore, Billerica, MA, USA) to remove bacterial cells, and storing at 4 °C. Phage titers were determined by plaque assay.

Phage DNA extraction was performed using the Lambda Mini Kit (Qiagen, Valencia, CA, USA) with the following modifications. Phage particles were precipitated in Buffer L2 by centrifugation at 20 000 *g* at 4 °C for 1 h. Extracted DNA was precipitated overnight at − 20 °C before centrifugation.

Phage electron microscopy
Copper grids (400 mesh formvar per carbon film) (Ted Pella, Redding, CA, USA) were glow discharged. Phage cultures were applied, followed by a wash with 0.22-μm-filtered water. The samples were stained with 1% uranyl acetate and examined under a JEOL JEM-1200EX electron microscope (JEOL, Peabody, MA, USA) with an accelerating voltage of 80 kV.

Phage genome sequencing, assembly and annotation
Phage genomes were sequenced in multiplex using the Roche GS FLX Titanium (Roche, Branford, CT, USA) or the Illumina MiSeq (Illumina, San Diego, CA, USA) platform. Sequence reads were initially assembled using MIRA 3.2.1 (Chevreux *et al.*, 1999), and the resulting contigs were manually finished in Consed 23.0 (Gordon *et al.*, 1998). Some phage genomes required additional PCRs and amplicon sequencing to fill the gaps between contigs. Fully assembled phage genomes were annotated using Genemark.hmm (Lukashin and Borodovsky, 1998) and Glimmer v3.02 (Delcher *et al.*, 1999) with manual corrections. All phage genome sequences have been deposited in GenBank under BioProject PRJNA173665 with accession numbers JX570702-JX570714, KJ578758-KJ578792.

Genome analysis and phylogenetic tree construction
Sequences present in all 62 phage genomes were defined as core regions of the phage genome. To identify these core regions, we first generated alignments between the PA6 genome and each of the other 61 phage genomes using Nucmer (Kurtz *et al.*, 2004). This yielded 61 sets of starting and ending coordinates describing intervals within the PA6 genome that align with a given phage genome. We then calculated the core regions for all phages by determining the overlapping intervals between all of the 61 coordinate sets. The core region sequences were concatenated for the subsequent multiple sequence alignments. Single nucleotide polymorphisms (SNPs) in the core regions were identified by using the 'show-snps' option of Nucmer with the default setting. In addition, the set of non-synonymous SNPs was obtained by masking the third codon positions in the coding regions. Using MEGA5 (Tamura *et al.*, 2011), phylogenetic trees were constructed by the Neighbor Joining method from p-distances based on all SNP sites in the core regions or only the non-synonymous SNPs. Bootstrapping was performed on 1000 replicates.

Analysis of nucleotide polymorphism within single phage strains
The genetic variation within a phage strain was measured by the number of SNPs found in the sequencing data from a clonal phage population. The SNPs were identified as sites in a strain's genome assembly that were covered by $\geqslant 30$ reads with Phred quality score $\geqslant 30$ and with $\geqslant 10\%$ of these reads differing from the consensus sequence.

Analysis of metagenomic shotgun sequencing data of the skin microbiota
Phage diversity was analyzed in the metagenomic shotgun sequencing data from 27 retroauricular crease samples collected in the HMP (The Human Microbiome Project Consortium, 2012). The SRA accessions of these samples are SRS013261, SRS024598, SRS013258, SRS024596, SRS019016, SRS019015, SRS019033, SRS019063, SRS019064, SRS019081, SRS024655, SRS024620, SRS020263, SRS020261, SRS017851, SRS017849, SRS057083, SRS024482, SRS045606, SRS058221, SRS018978, SRS058182, SRS016944, SRS046688, SRS015381, SRS052988 and SRS019116. Access to the phenotype data for this study (phs000228.v3.p1) was obtained from dbGaP. MIRA (Chevreux *et al.*, 1999) was used to identify phage reads in each data set by mapping against the *P. acnes* phage PA6 genome. Parameters similar to the default were used: $\geqslant 20$ nt overlap and $\geqslant 60\%$ identity. To estimate the number of *P. acnes* phage strains in each data set, we first performed a *de novo* assembly using the extracted phage reads from the metagenomic data. We then aligned all the phage reads to the resulting contigs to identify SNPs in the core genome regions. The assemblies and alignments were manually inspected using Consed (Gordon *et al.*, 1998). The same criteria used for nucleotide polymorphism identification in single phage strains were applied as described above.

Analysis of phage genes under diversification
Multiple sequence alignments of Group VI and Group VIII phages and their related phages (PHL037M02 and PHL073M02) were generated using MAFFT (Katoh *et al.*, 2002). The positions of all mismatches and gaps were recorded. Sites of discrepancy were plotted in Artemis (Rutherford *et al.*, 2000).

Propionibacterium culture
P. acnes, *P. humerusii*, *P. granulosum* and *P. avidum* strains were cultured under anaerobic conditions in Clostridial Reinforced medium (Oxoid, Thermo Fisher Scientific, Waltham, MA, USA) at 37 °C for 4–6 days. *Propionibacterium* cultures were used to prepare top agar overlays for phage culture on A-media plates.

Phage resistance test
The resistance/susceptibility of *Propionibacterium* strains against phages was determined using a modified cross-streak assay. Fifteen of the 48 newly sequenced phages were randomly chosen for the analysis. Two sets of phages that each belongs to the same group, PHL010M04 and PHL066M04 in Group VIII and PHL115M02, PHL085N00, and PHL085M01 in Group VI, were included. The bacterial strains were streaked across in A-media plates, along with ATCC6919 on the same plate as a control. Approximately 5 µl of 10^6 PFU ml^{-1} phage suspension was spotted onto each bacterial streak. The plates were incubated at 37 °C anaerobically for 2 days. At least five replicates of each cross-streak experiment were performed to determine whether the strains were susceptible or resistant.

For the strains that showed resistance in the modified cross-streak experiment, we further quantitatively determined the resistance by assaying the efficiency of plaquing of the phages relative to *P. acnes* strain ATCC6919, calculated as the following:

$$\text{Resistance} = \frac{1}{\text{Efficiency of plaquing}} = \frac{\text{Titer of phage strain } X \text{ on ATCC6919}}{\text{Titer of phage strain } X \text{ on bacterial strain } Y}$$

We considered a 100-fold or greater increase in efficiency of plaquing to be evidence of resistance.

The plaques on cross-streak plates were visually inspected by one person and scored for turbidity based on the re-growth of the bacteria after plaque formation using the following scale: 0 = clear, 1 = little to no re-growth, 2 = mild re-growth, 3 = moderate re-growth and 4 = heavy re-growth. The average plaque turbidity score of all the strains from the same *P. acnes* clade was calculated for each of the tested phages and was compared among different clades.

Pseudolysogeny characterization
PCR was performed on phage suspensions using the primers annealing to the ends of the phage genomes (Forward 5′-CCGAAGCCGACCACATCACACC-3′, Reverse 5′-TCATCCAACACCTGCTGCTGCC-3′) to determine whether phage genomes are circularized. The PCR was run under the following conditions: initial denaturation at 94 °C for 5 min, 35 cycles of denaturation at 94 °C for 45 s, annealing at 53 °C for 35 s and extension at 72 °C for 1 min, with a final extension at 72 °C for 10 min.

P. acnes ATCC6919 cultures, which were re-grown after lytic infection with specific phages, were tested for superinfection immunity by passaging sequentially two to four times without further phage infection. Phage resistance was assayed using the same cross-streak method described above. The presence of phage DNA in re-grown cultures was determined by PCR using the primers targeting the phage *gp11* gene (Forward 5′-GGCTGGAACACGTAAAGCG-3′, Reverse 5′-CACGATCGATCAACTCAACC-3′). The PCR was run under the following conditions: initial denaturation at 94 °C for 5 min, 35 cycles of denaturation at 95 °C for 45 s, annealing at 58 °C for 35 s and extension 72 °C for 1 min, with a final extension at 72 °C for 10 min.

CRISPR analysis
CRISPR spacer sequences were previously identified in *P. acnes* genomes (Fitz-Gibbon *et al*., 2013; Tomida *et al*., 2013). The spacer sequences were aligned against all phage genomes using BLASTn. Protospacers with up to two mismatches were identified.

Results

Phage isolation and genome features
In an effort to determine the diversity of the skin microbiota, we previously collected 203 skin samples from 179 individuals, including 94 samples from acne patients and 109 samples from healthy individuals with clear skin (Li, 2010; Fitz-Gibbon *et al*., 2013). Twenty-four individuals were sampled twice over a 4–6 month period. When we cultured the skin samples for bacteria, we observed phage plaques in 49 of these samples: 14 from acne patients and 35 from healthy individuals. *P. acnes* phages were found more frequently in samples from healthy individuals than from acne patients. Our rate of phage detection (24% of investigated samples) is similar to those previously reported, ranging from 26 to 30% (Marples *et al*., 1973; Puhvel and Amirian, 1979). The phages that we isolated have the morphology of siphoviruses as previously described (Farrar *et al*., 2007). A representative electron micrograph is shown in Supplementary Figure S1. Among the phage isolates obtained from these samples, we selected 21 phages from acne patients and 27 from healthy individuals for whole genome sequencing using 454 or MiSeq platforms (Supplementary Table S1). In some samples, multiple phage plaques were isolated and selected for sequencing. All phage genomes were assembled, completed and annotated (Supplementary Figure S2). A representative phage genome is shown in Figure 1.

The *P. acnes* phage genomes are highly similar to each other (Supplementary Figure S2). The 48 phage genomes have comparable sizes (29.0–29.8 Kb) and GC contents (53.7–54.5%) (Supplementary Table

S1). The sequence identity between any pair of genomes ranges from 85.2 to 100%. On average 45 open reading frames were predicted in each genome. Consistent with previous reports (Farrar *et al.*, 2007; Lood and Collin, 2011; Marinelli *et al.*, 2012), these open reading frames were arranged compactly within the left and right arm regions of the genome (Figure 1, Supplementary Figure S2). Our analysis of the 48 new phage genomes supports the annotation of the *gp22/gp23* locus as a single open reading frame (495 to 522 bp) on the minus strand (Supplementary Figure S2). This is different from previous annotations based on a small number of genomes (Farrar *et al.*, 2007; Lood and Collin, 2011).

Phylogenetic relationships among the phage genomes

To determine the genome diversity of *P. acnes* phages, we compared 62 sequenced phage genomes, including our 48 phage genomes and 14 previously published genomes (Farrar *et al.*, 2007; Lood and Collin, 2011; Marinelli *et al.*, 2012). Similar to their bacterial host, *P. acnes* phages have limited genomic diversity. All 62 phages are highly similar in genome sequence. The core regions, which are shared by all sequenced genomes, consist of 22 348 bp (76% of the average genome length) and contain 7232 SNPs. The average distance among the phages was 0.257 (substitution rate at the SNP sites). A phylogenetic tree constructed based on these SNPs (Figure 2), or only the non-synonymous SNPs (Supplementary Figure S3), shows that no particular phylogenetic clades were found among the phages.

Despite the lack of phylogenetic lineages among the *P. acnes* phage genomes, we observed several groups (I–IX), each of which consists of nearly identical phages (Figure 2). The phages within each group (I–IX) differ by no greater than 14 bp within the entire 29 kb genome (Supplementary Table S2). As a comparison, the average pairwise difference among all 62 phages is 3176 bp. To determine whether the nearly identical phages within each group represent clones of the same phage strain, we estimated the frequency of genetic polymorphism in each of our phage isolates. We mapped all available sequence reads of each phage to its assembled consensus genome, including the sequence reads not used in the genome assembly process. We identified the nucleotide positions with a minor allele frequency $\geqslant$10% and covered by at least 30 reads with Phred quality $\geqslant$30. We found that each phage genome contains 0–11 polymorphic sites (Supplementary Table S1). The number of polymorphisms in each assembled genome did not increase beyond 11 sites despite the large numbers of sequencing reads obtained (up to 9120× genome coverage). From this analysis, we conclude that the background level of genetic polymorphism within a clonal *P. acnes* phage isolate is ~11 bp. Thus, phage isolates with a similar or smaller number of nucleotide differences throughout the entire genome can be considered as belonging to the same phage strain. Since the phages within each group (I–IX) differ by at most 14 bp (Supplementary Table S2), they likely represent clones of the same phage strain.

Diversity of P. acnes *phages in the human skin*

The relationships among the phages within each group (I–IX) provide insights on the diversity of *P. acnes* phages in the skin microbiota. We observed three types of relationships among the highly similar phages within the groups (Figure 2 and Supplementary Table S2). First, phages within the same group were isolated from the same sample of the same individual. These include PHL067M01, PHL067M09 and PHL067M10 in Group I; PHL082M00, PHL082M02, PHL082M03 and PHL082M04 in Group II; PHL064M01 and PHL064M02 in Group V and PHL116M00 and PHL116M10 in Group IX. Only one other pair of phages, PHL117M00 and PHL117M01, which were isolated from the same sample, was not the same strain. Our data suggest that while an individual microbiota can harbor multiple strains of phages, it is likely more common that one strain of *P. acnes* phage dominates the phage population. Second, phages within the same group were isolated from different samples of the same individuals over a period of 14 to 21 weeks. These include PHL085M01 and PHL085N00 in Group VI; PHL114L00 and PHL114N00 in Group VII and PHL151M00 and PHL151N00 in Group VIII. All paired phages from our longitudinal samples were nearly identical to each other. This suggests that the same phage strain can persist in an individual skin microbiota. Third, some phages within the same group were isolated from different individuals, such as the phages in Groups III, IV, V, VI and VIII. Of the 43 unique phage strains represented by all 62 isolates sequenced to date, 5 strains from our study currently show evidence of inhabiting more than one individual. This suggests that a pool of common *P. acnes* phage

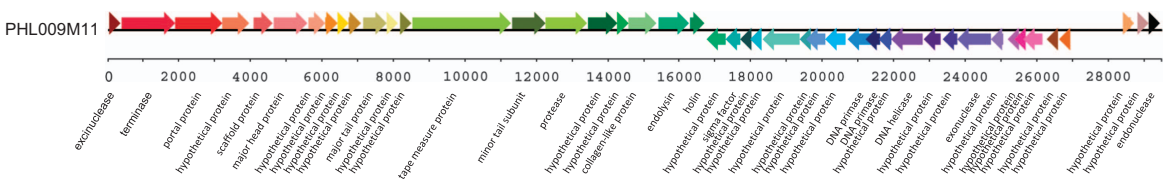


Figure 1 A representative genome of the newly sequenced *P. acnes* phages. The annotated genome of PHL009M11 is shown as a representative of the 48 newly sequenced *P. acnes* phage genomes. On average 45 open reading frames are encoded in each phage genome.

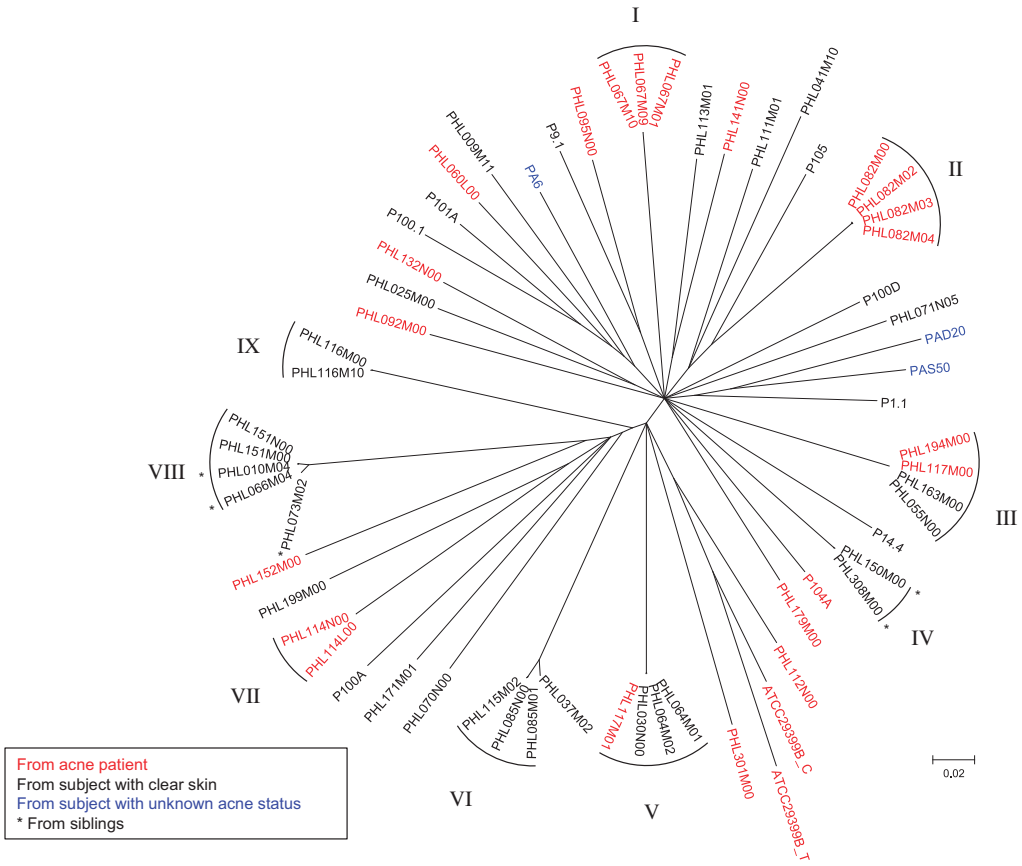


Figure 2 *P. acnes* phages are highly similar to each other with no significant phylogenetic lineages observed. A phylogenetic tree of the 62 currently sequenced phage genomes was constructed based on the 7232 SNPs in the core regions. No significant lineages were observed. Nine groups of nearly identical phages are indicated, highlighting that the phages within each group belong to the same strain. Branches with bootstrap values <80 (based on 1000 resamplings) were collapsed.

strains exists in the human population. Interestingly, among the five shared strains, two inhabited related individuals. In Group IV, the two nearly identical phages, PHL150M00 and PHL308M00, were isolated from two brothers. Two of the four phages in Group VIII, PHL010M04 and PHL066M04, and their closely related phage strain PHL073M02 were isolated from three siblings (Figure 2 and Supplementary Table S2). This suggests that transmissions of skin bacteriophages, either directly or via the transmission of phage-carrying bacterial hosts, are likely to occur between related individuals.

P. acnes *phage populations in the skin microbiome*

To validate that our above findings on the phage diversity in the human skin microbiota were not biased due to isolated phages, we analyzed the skin metagenomic shotgun sequencing data collected from healthy individuals in the HMP (The Human Microbiome Project Consortium, 2012). Although *P. acnes* phages were previously found in metagenomic sequencing data, their diversity was not analyzed. This is the first time that the population diversity of *P. acnes* phages in the skin microbiome is characterized. Among the available 27 HMP skin samples, 9 were collected from the left retroauricular crease and 18 were collected from the right retroauricular crease (Supplementary Table S3). Some samples were collected from the same individuals. We first extracted the *P. acnes* phage reads from each sample. The number of phage reads was from 24–612 512 (0–2060 × coverage on PA6 genome, Supplementary Table S3), independent of the sequencing depth of each sample, suggesting that the relative abundance of *P. acnes* phages in the skin microbiota can vary in a large range among individuals.

To determine *P. acnes* phage populations in individual skin microbiota, we next assembled *P. acnes* phage genomes in each metagenomic data set. Seven samples (25.9% of total samples) had a modest to high sequencing coverage of *P. acnes* phages (17 × –2060 ×) (Supplementary Table S3), which allowed metagenomic assembly. We

performed the SNP analysis on the assembled phage genomes using the same criteria as in the analysis of complete phage genomes described above. To evaluate the effect of sequencing depth on the detection rate of SNPs, for samples with >30× phage coverage, we repeated the phage genome assembly and SNP analysis using only portions of phage reads at various coverages (20×, 100×, 250×, 500×, 1000× and 1500×), as applicable. One of the samples, HMP20, had >1000 SNPs in the core regions of the assembled phage genome, which leveled off to 1353 sites when all the sequence reads (538×) were used in the assembly (Figure 3, Supplementary Table S4). This suggests that the *P. acnes* phage pool in HMP20 was adequately sampled and that it likely consists of two dominant phage strains based on phage genome comparison (Supplementary Materials).

On the other hand, three other samples, HMP04, HMP09 and HMP15, each had six or fewer SNPs in the core genome regions, despite having substantially higher sequencing coverages than HMP20. This suggests that these samples each harbor only one dominant *P. acnes* phage strain. This result supports our earlier conclusion that an individual skin microbiota is often dominated by one *P. acnes* phage strain. We were able to assemble high-quality draft genomes of *P. acnes* phages from these three metagenomic data sets, which cover 96.4–96.7% of the core genome regions. They are typical of the *P. acnes* phage strains, with high similarities to the 62 sequenced genomes. A phylogenetic tree including these three new genomes with the 62 isolated phage genomes is shown in Supplementary Figure S4. Samples HMP04 and HMP09 were collected from the left and right retroauricular crease of the same individual, and the phage genomes assembled from these two samples are highly similar, potentially originating from the same strain.

The assembled phage genomes in the remaining three samples, HMP03, HMP08 and HMP24, contained <20 SNPs, which are in the range of the SNPs found in a single phage strain. Due to their lower phage coverages, it is yet inconclusive whether single or multiple dominant phage strains were present in these communities.

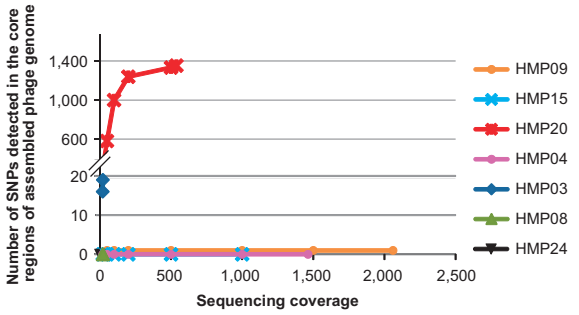


Figure 3 *P. acnes* phage population in the skin microbiota is often dominated by a single strain. The numbers of SNPs identified in the core regions of *P. acnes* phage genomes are shown at different sequencing coverages. The phage genomes were assembled from the HMP metagenomic shotgun sequencing data.

Our analysis of the phage diversity from the HMP metagenomic data showed that among the skin communities that had detectable *P. acnes* phages, three harbored only one dominant strain while one harbored two different strains. These findings are consistent with our conclusions based on the isolated phages from our study cohort.

Phage genes under diversification

Phylogenetically related strains can reflect phage diversification under selection. Two phages, PHL037M02 and PHL073M02, are highly related to the members of Groups VI and VIII, respectively (Figure 2). However, they contain many more nucleotide variations than 11 bp, and thus are considered separate strains based on our criterion. Nonetheless, their high degree of similarity to the members of these groups may reflect recent selective pressures driving phage diversification. We identified 160 sites of nucleotide variations between PHL037M02 and the Group VI members, 50 of which are non-synonymous. All but one of these genetic differences is located in a region encoding Gp16, Gp17 and Gp18, as annotated in the genome of phage PA6 (Supplementary Figure S5A). The exact functions of these genes are unknown, but their location near the 3' end of the left arm between structural protein genes and lysis protein genes suggests that they could encode late-acting proteins. Based on the McDonald–Kreitman test (Egea *et al.*, 2008), gene *gp17* is under selection ($P = 0.011$).

We identified 81 sites of nucleotide variations between PHL073M02 and the Group VIII members. The sequence differences lie primarily within the region encoding an endolysin and a putative type II holin (Gp20 and Gp21, Supplementary Figure S5B). These lytic cycle proteins permeabilize the cell membrane and degrade the extracellular peptidoglycan layer to release new phage particles from the bacterial host. As previously mentioned, PHL073M02 and two Group VIII members, PHL010M04 and PHL066M04, were isolated from three siblings living in the same household. Thus, they likely originated from the same ancestral phage strain. This suggests that the endolysin and holin genes, which are essential for phage multiplication, are under rapid evolution. Consistent with this finding, in all sequenced phages, we found frequent amino acid variations in these two gene products. This suggests that mechanisms determining host bacterium lysis specificity and kinetics may be under selection in these phages. The sequence variation sites in these two genes among the phages may be potential targets for phage engineering to manipulate their lytic activities against bacterial hosts.

The two highly similar phage genomes assembled from the HMP samples, HMP04 and HMP09, which were collected from the same individual, differed by

222 nucleotides. Most of these variations are centered at the 5′ end of the right arm of the genome, which encodes putative regulatory or DNA-binding proteins of largely unknown functions.

Range and specificity of Propionibacteria–phage interactions

To determine whether *P. acnes* phages modulate the relative abundances of different *P. acnes* strains in the skin microbiota by selective killing, we characterized the host range and specificity of *P. acnes* phages. We tested 15 of the 48 sequenced phages against a panel of 74 *Propionibacterium* strains, including 67 *P. acnes* strains, 3 *P. humerusii* strains, 1 *P. granulosum* strain and 3 *P. avidum* strains. Except for the *P. acnes* strains KPA171202, ATCC11828, HL201PA1 and HL202PA1, all of these *Propionibacterium* strains were isolated from the same cohort of subjects sampled for phages. The genomes of all 67 *P. acnes* strains and 3 *P. humerusii* strains have been sequenced (Fitz-Gibbon *et al.*, 2013; Tomida *et al.*, 2013). Our bacterial collection included all major lineages of *P. acnes* found on the human skin, with multiple strains representing each of the major clades, IA-1, IA-2, IB-1, IB-2, IB-3 and II, as well as one type III strain. We constructed a phylogenetic tree of the 67 *P. acnes* strains based on the SNPs in their core genomic regions (Figure 4a) (Tomida *et al.*, 2013). Using a modified cross-streak method, we determined the resistance/susceptibility of each of the 74 bacterial strains against the 15 phages. In total, 1110 bacterium–phage interactions were measured. Each experiment was repeated a minimum of five times. For the bacterial strains that showed resistance to phages, we determined the fold increase in resistance by measuring efficiency of plaquing (EOP) relative to the *P. acnes* strain ATCC6919, which is known to be susceptible to all tested phages.

We found that the outcome of the bacterium–phage interactions is *P. acnes* lineage dependent. All type I *P. acnes* strains (clades IA-1, IA-2, IB-1 and IB-2) except clade IB-3 were susceptible to all tested phages (Figure 4a). Among them, the phages often formed turbid plaques on *P. acnes* strains of clade IA-1, but clear plaques on strains of clades IB-1 and IB-2, as summarized in Figure 5. This suggests that these phages engage in two different states depending on the host strains: a pseudolysogenic response in clade IA-1 strains, and a lytic cycle in clade IB-1 and IB-2 strains.

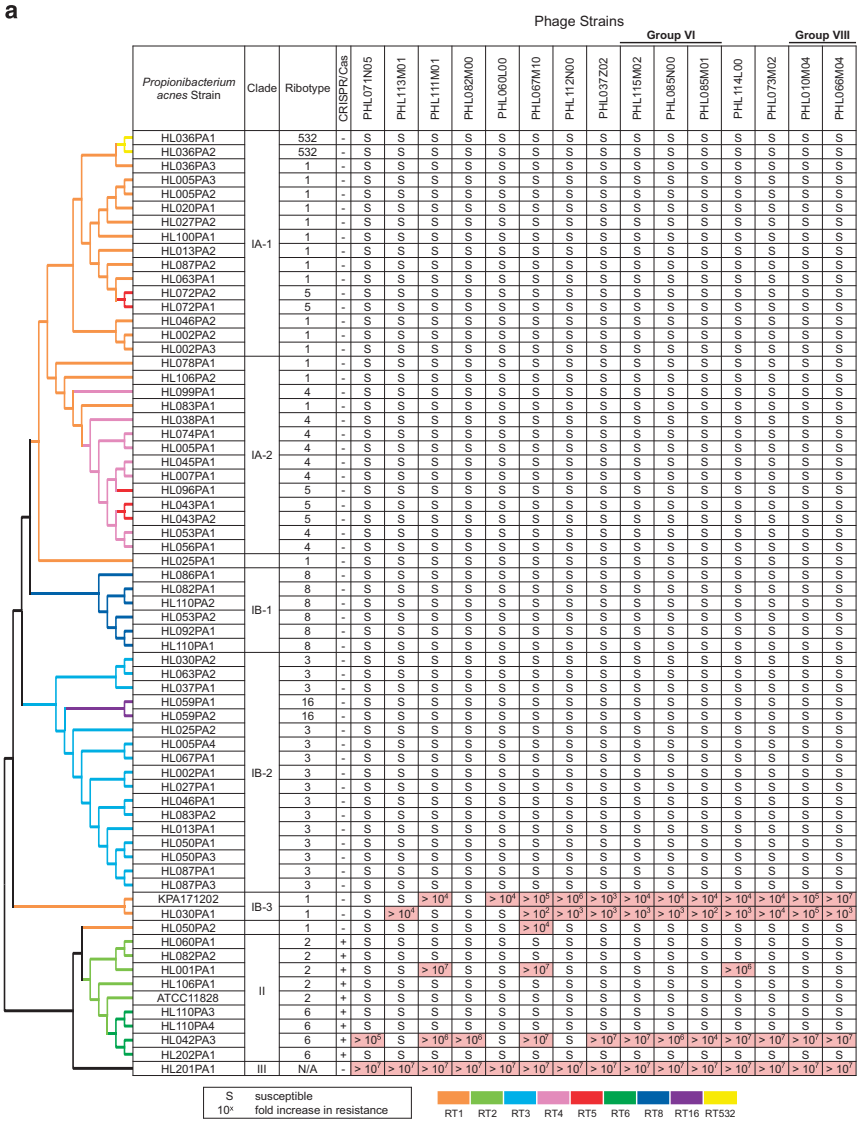
Certain *P. acnes* strains of clades IB-3, II and III are highly resistant to multiple phages. Two strains of clade IB-3 (KPA171202 and HL030PA1) were highly resistant to most of the tested phages with a $\geqslant$100-fold increase in resistance. The genomes of clade IB-3 show that these strains encode components of a restriction modification system (genes PPA1611 and PPA1612 in KPA171202). This may explain their resistance to phages. Among the nine type II strains, two strains, HL001PA1 and HL042PA3, were highly resistant to some of the phages. This is consistent with previous observations that strains of this type were more frequently resistant to phages (Webster and Cummins, 1978). The resistance to phages observed in type II strains could be partially attributed to the CRISPR mechanism encoded in their genomes, which is addressed below. The only type III strain tested, HL201PA1, was resistant to all 15 phages. Since type III strains are not commonly found on the skin of the face, it is possible that these *P. acnes* phages isolated from the face have not yet evolved a mechanism to infect type III strains.

To determine whether *P. acnes* phages modulate the abundance and diversity of other species in addition to *P. acnes* in the skin microbiota, we investigated the host range of *P. acnes* phages in related species. Strains of other human skin-associated *Propionibacteria*, including 3 strains of *P. humerusii*, 1 strain of *P. granulosum* and 3 strains of *P. avidum,* were tested against the 15 phages (Figure 4b). *P. humerusii* is a newly defined species (Butler-Wu *et al.*, 2011). In our previous study, *P. humerusii* was one of the major species found on the skin with a relative abundance of 1.9% in the pilosebaceous unit based on 16S ribosomal RNA analysis (Fitz-Gibbon *et al.*, 2013). It is closely related to *P. acnes* with $>$98% identity in the 16S ribosomal RNA gene sequence. *P. granulosum* and *P. avidum* are common skin commensals (Cummins, 1976; Ördögh and Hunyadkürti, 2013). While all tested *P. granulosum* and *P. avidum* strains showed strong resistance to all the phages, two *P. humerusii* strains, HL037PA2 and HL037PA3, were susceptible to all the phages tested. The third *P. humerusii* strain, HL044PA1, was susceptible to 10 of the 15 phages tested. Our results show that the host range of *P. acnes* phages is not limited to *P. acnes* but also includes a closely related *Propionibacterium* species, suggesting that *P. acnes* phages may also be able to modulate *P. humerusii* populations in the skin microbiota.

P. acnes *phages can adopt a pseudolysogenic state depending on the host* P. acnes *strains*

It has been suggested that *P. acnes* phages may enter pseudolysogeny as an alternative to the lytic cycle (Farrar *et al.*, 2007; Lood and Collin, 2011). As described above, we discovered that *P. acnes* phages often adopt a pseudolysogenic state in clade IA-1 strains, but rarely in clade IB-1 and IB-2 strains, suggesting that their pseudolysogeny is dependent on the host *P. acnes* strains (Figure 5). In support of the existence of pseudolysogeny in *P. acnes* phages and consistent with the result by Marinelli *et al.*, (2012), our genome sequencing data revealed the ends of the phage genomes to be flanked by 11-nucleotide single-stranded overhangs. Previous reports suggested that these overhangs may be involved in the circularization of the phage genomes (Farrar *et al.*, 2007; Lood and Collin, 2011). To test

a



b

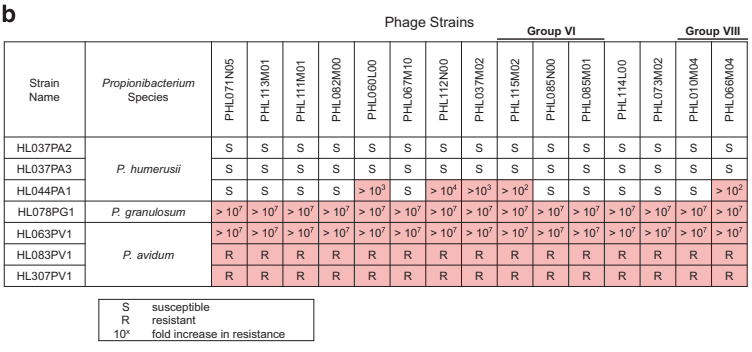


Figure 4 Host range and specificity of *P. acnes* phages. Resistance/susceptibility of *Propionibacterium* strains (in rows) against 15 of the 48 sequenced phages (in columns) is shown. (**a**) Phage infection outcomes of 67 *P. acnes* strains. All *P. acnes* strains in clades IA-1, IA-2, IB-1 and IB-2 were susceptible to the tested phages, while phage resistance was found in strains of clades IB-3, II and III, colored in pink. The dendrogram to the left shows the phylogenetic clades of *P. acnes* strains (Fitz-Gibbon *et al*., 2013). Only topology is shown. (**b**) Phage infection outcomes of three *P. humerusii* strains, one *P. granulosum* strain and three *P. avidum* strains show that *P. acnes* phages could infect and lyse *P. humerusii* strains, while *P. granulosum* and *P. avidum* were resistant to all tested phages.

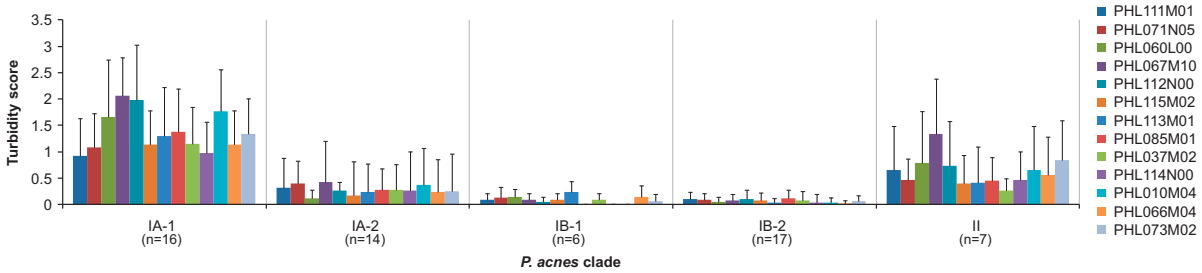


Figure 5 The frequency of phage pseudolysogeny varies among different *P. acnes* strains. The turbidity of the phage plaques formed on *P. acnes* culture plates was examined. Data from 60 *P. acnes* strains challenged by 13 phages were recorded and summarized. The 60 *P. acnes* strains belong to five clades: IA-1 ($n=16$), IA-2 ($n=14$), IB-1 ($n=6$), IB-2 ($n=17$) and II ($n=7$). The average turbidity score of the bacterial strains in each clade challenged by each phage (in colors) is shown. Independent of the phages tested, the plaque turbidity score varied among different *P. acnes* strains: it is high in clade IA-1 and low in clades IB-1 and IB-2. This suggests that the phages tend to enter a pseudolysogenic state in clade IA-1 strains and a lytic cycle in clades IB-1 and IB-2 strains.

this hypothesis, we designed PCR primers annealing to the ends of the phage genomes. A PCR product spanning the two ends with the predicted size (~735 bp) and sequence was amplified from all the phages tested, suggesting that the phage DNA can exist in a circular form mediated by the overhangs at the ends (Supplementary Figure S6A).

To demonstrate the pseudolysogenic properties of *P. acnes* phages, we infected *P. acnes* strain ATCC6919, a clade IA-1 strain (Liu *et al.*, 2014), with five different phages (PHL060L00, PHL112N00, PHL037M02, PHL073M02 and PHL114L00). Following the formation of plaques by each phage, re-growth of the bacteria was observed starting in the center of the plaque regions. The re-grown bacteria showed no evident lysis when challenged by 13 phages that previously could lyse this *P. acnes* strain, including the phages of the initial exposure (Supplementary Figure S6B). This suggests that the bacterial host gained superinfection immunity after infection by these phages.

To further determine whether the phage DNA exists as an episome in the bacterial host after infection, we tested the presence of phage DNA in four distinct colonies of the re-grown ATCC6919 culture that was initially infected with PHL060L00 and subsequently passaged. Two of the four colonies produced an expected 437 bp amplicon in a PCR targeting the *gp11* gene (Supplementary Figure S6C), supporting the presence of phage in a subpopulation of the re-grown *P. acnes* as an episome. In addition, we sequenced the genomic DNA extracted from two ATCC6919 cultures, each of which was passaged from a re-grown culture after phage infection. In both cases, the phage reads obtained from whole genome sequencing were assembled into complete genomes that are separate from host *P. acnes* contigs, supporting earlier evidence that *P. acnes* phage DNA does not integrate into the host genome (Farrar *et al.*, 2007; Lood and Collin, 2011), but exists as an extrachromosomal element.

In summary, all above results suggest that *P. acnes* phages can adopt a pseudolysogenic state in clade IA-1 strains.

Resistance to bacteriophages does not correlate with the presence of matching CRISPR spacers in type II P. acnes *strains*

Since certain type II *P. acnes* strains are resistant to phages, we next investigated whether the CRISPR/Cas mechanism could explain the resistance of the type II strains against phages. Among the 67 *P. acnes* strains, 9 belong to type II and encode CRISPR/Cas elements. Each of the type II strains has two to nine 33-bp spacers in their CRISPR arrays (Tomida *et al.*, 2013). In total, they encode 44 spacers, 28 of which are unique. We identified 42 unique protospacers in the 15 tested phage genomes that match any of the 28 unique spacer sequences in the 9 type II *P. acnes* strains. Because the CRISPR/Cas system has been shown to tolerate a limited number of mutations in protospacer targets (Semenova *et al.*, 2011; Manica *et al.*, 2013), in our analysis we allowed up to two mismatches for a sequence to be considered a recognizable protospacer. Similar results were obtained when only perfectly matching protospacers were considered. We found that all identified protospacers are located primarily on the left arm of the phage genomes, which is more conserved than the right arm (Supplementary Figure S7). In addition, the locations of the protospacers are generally conserved among all phage genomes that harbor the same protospacer sequences. These suggest that the CRISPR/Cas system tends to target the more conserved regions of the phage genomes.

In contrast to a prior report (Marinelli *et al.*, 2012), we found that the resistance/susceptibility of the nine type II *P. acnes* strains against phages did not correlate with the presence/absence of at least one phage-matching CRISPR spacer ($r=0.234$, Figure 6). There were multiple observations that even though the strain encodes a matching CRISPR spacer, it was susceptible to the corresponding phage. To determine whether more than one matching spacer is needed to confer phage resistance, we analyzed the correlation between the resistance/susceptibility and the number of matching spacers encoded in the CRISPR array. No strong correlation ($r=0.26$) was found, suggesting that encoding multiple matching spacers does not

necessarily protect the strain from phage infection. In addition, no presence of particular spacer sequences can explain the resistance/susceptibility to phages. When we allowed no mismatches or up to five mismatches for a sequence to be considered a recognizable protospacer, the correlations were not much improved. This suggests that having matching CRISPR spacers is insufficient for these strains to be resistant to phages and that other regulation mechanisms are likely to be involved.

Phages can escape the CRISPR/Cas mechanism through mutations at sites involved in protospacer recognition. The short nucleotide motif adjacent to the protospacer, known as the protospacer-adjacent motif (PAM), is highly conserved among targets of CRISPR/Cas systems (Mojica *et al.*, 2009). Mutations in these nucleotides have been found to disrupt CRISPR-mediated resistance despite complete identity in the protospacer sequence (Semenova *et al.*, 2011, 2009; Westra *et al.*, 2012).

To determine whether the lack of correlation between bacterial resistance and the presence of matching spacer sequences is due to mutations in the PAM sequence, we examined the PAM regions of the 137 protospacers identified in the phages from this study and from the study by Marinelli *et al.*, (2012). The protospacers can be separated into two groups depending on the outcomes of the bacterium–phage interaction—resistant or susceptible to the phages (Figure 7). Fifty-three protospacers belong to the first group, where the bacterial hosts with matching spacer sequences were resistant to the phages. The other 84 protospacers belong to the second group, where lysis of the bacterial hosts occurred despite the presence of a matching spacer. Among the first group of protospacers, we observed a conserved 'AA' motif immediately upstream of the protospacer sequence (Figure 7a). The AA motif is also conserved in the PAM sequences of the 84 protospacers from the second group where the bacterial hosts were susceptible to the phages, except in 3 cases, where an A-to-G mutation occurred. These three mutated positions were found in phages PHL060L00 and PHL112N00, which are targeted by spacers encoded in *P. acnes* strain HL042PA3, and in phage PHL113M00, which is targeted by a spacer encoded in strain HL110PA3 (Figure 7b). However, these mutations in the PAM sequence cannot explain the lack of resistance against these phages, as all three phages contain one to three additional protospacers that are recognizable by the spacers encoded in the corresponding bacterial strains. All of these additional protospacers have intact PAMs. Our findings are corroborated by the previous observation of unexplained susceptibility to *P. acnes* phage despite the presence of matching spacers (Marinelli *et al.*, 2012). In summary, our data demonstrate that encoding CRISPR spacers that match against the genome of an invading *P. acnes* phage is often insufficient for an effective defense in *P. acnes* and that other regulation mechanisms are likely to be involved.

Discussion

The skin microbial community has an essential role in defending against the invasion and colonization of pathogens. Although the bacterial and fungal compositions of human skin have been characterized at multiple sites in healthy individuals (The Human Microbiome Project Consortium, 2012; Gao *et al.*, 2007; Paulino *et al.*, 2008; Grice *et al.*, 2009; Gao *et al.*, 2010; Ursell *et al.*, 2012; Blaser *et al.*, 2013; Findley *et al.*, 2013; Fitz-Gibbon *et al.*, 2013; Nakatsuji *et al.*, 2013), we have limited understanding of the skin virome (Foulongne *et al.*, 2012; Ma *et al.*, 2014). Whether bacteriophages have a role

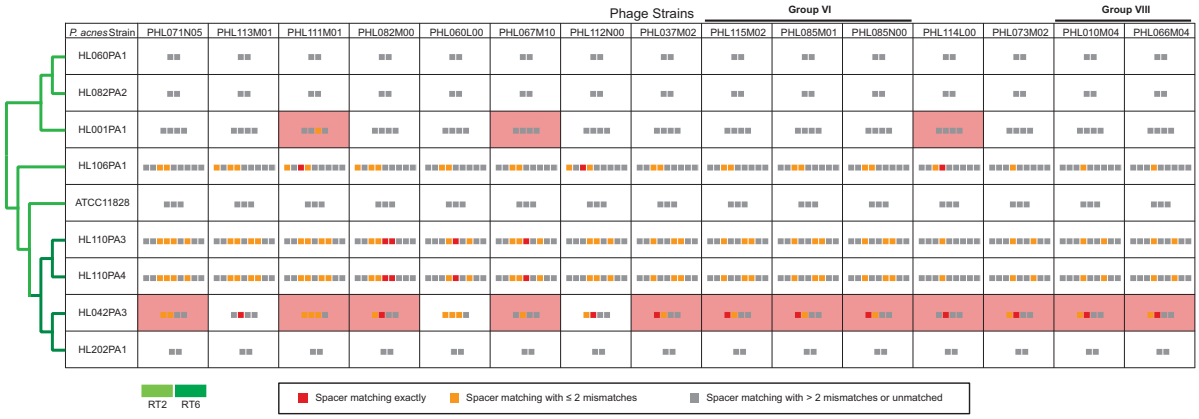


Figure 6 Phage resistance is not correlated with the presence of matching CRISPR spacers in type II *P. acnes* strains. The pattern of resistance/susceptibility of the 9 type II *P. acnes* strains against 15 phages did not correlate with the presence/absence of matching CRISPR spacers, that is, the pink cells do not perfectly overlap with the cells containing red or orange pixels. *P. acnes* type II strains are shown in rows and the phages are shown in columns. Pink cells indicate that the bacterial strains are resistant to the corresponding phages. The colored pixels in each cell represent the CRISPR spacers encoded in each *P. acnes* strain. Each red pixel indicates that this spacer has an exact protospacer match in the corresponding phage. Each orange pixel indicates that this spacer has a partially matched protospacer (with one to two mismatches) in the corresponding phage. Gray pixels indicate no matched protospacers found in the corresponding phage.

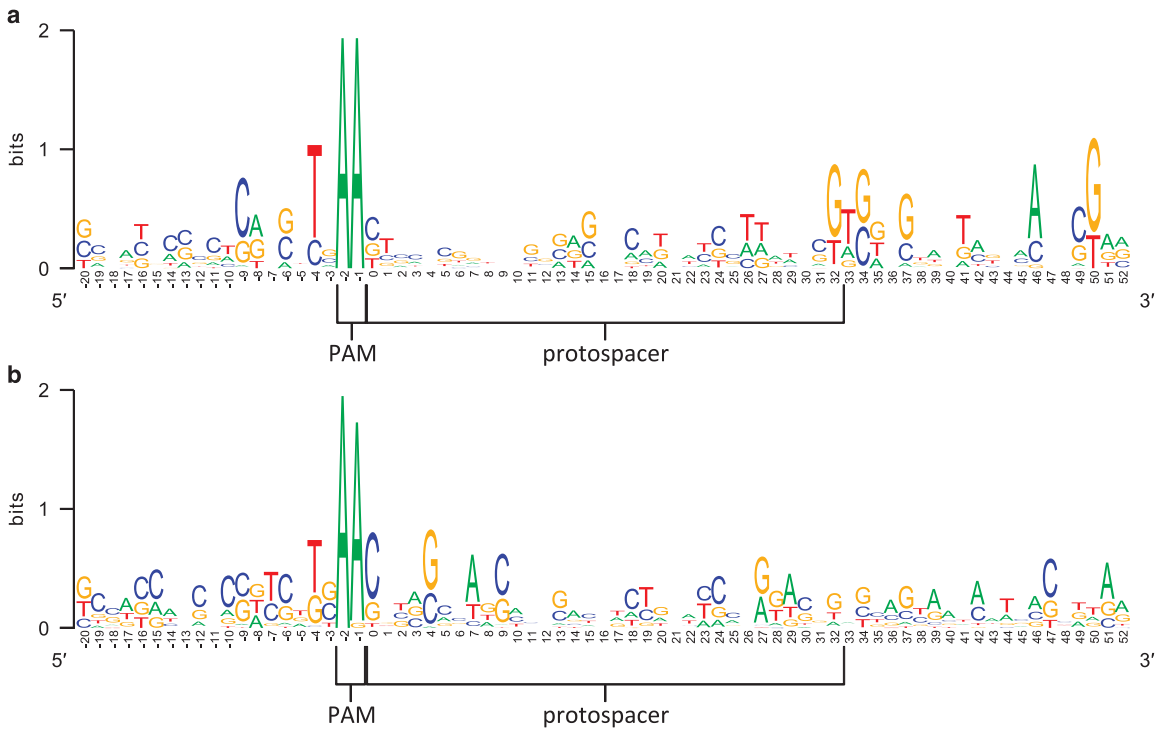


Figure 7 Sequence conservation in protospacers and PAMs. A logo representation of the protospacer sequences that match *P. acnes* CRISPR spacers (within two mismatches) is shown. The PAM sequence of the protospacers is also shown. (**a**) Sequence conservation among the protospacer motifs from the phages unable to lyse the corresponding bacterial strains. (**b**) Sequence conservation among the protospacer motifs from the phages able to lyse the corresponding bacterial strains. The PAM sequences are conserved in both cases.

in skin health and disease by modulating the diversity and dynamics of the bacterial community on the skin is poorly understood.

The lack of knowledge of the skin virome is partly due to the limited microbial biomass of skin samples, which poses challenges in determining the composition and diversity of skin viruses. As methods are being developed in recent years for high-throughput sequencing from a limited amount of DNA (Duhaime and Sullivan, 2012; Duhaime *et al.*, 2012; Solonenko *et al.*, 2013) and for capturing metagenomic viral sequences in metagenomic sequencing data (DeLong *et al.*, 2006; Mizuno *et al.*, 2013), it is becoming feasible to study lowly abundant viral communities directly from metagenomic samples without enriching the viral portion. In this study, using genome sequencing and metagenomic analysis of the skin microbiome, we characterized a large number of *P. acnes* bacteriophages and their genome diversity in the skin of healthy individuals and acne patients. This effort significantly increased our knowledge about this viral population and allowed us to investigate its population diversity in the human skin microbiota.

Unlike phages of other species (Nolan *et al.*, 2006; Dorscht *et al.*, 2009; Sepúlveda-Robles *et al.*, 2012), *P. acnes* phages show a high genetic similarity. The phages may have evolved recently under limited selection, as no clear lineages can be found among the phages studied, suggesting that these phages may be relatively new to their hosts (Marinelli *et al.*, 2012). Alternatively, the high similarity among the phages may be attributed to the low genetic diversity of their host strains, as *P. acnes* has limited genome expansion among different lineages (Tomida *et al.*, 2013).

By comparing the phages isolated from different individuals (Figure 2), we found that in individuals harboring *P. acnes* phages, the skin community is often dominated by one *P. acnes* phage strain. Our metagenomic data analysis of the skin samples from the HMP cohort confirmed this finding (Figure 3), despite the differences between the two analyses in methodology, study cohorts and skin sites that were sampled. We also observed the same phages in the skin microbiota when the individuals were sampled 14–21 weeks later (Supplementary Table S2). Our findings are consistent with the intra-individual temporal stability found in the human oral and gut viromes (Pride *et al.*, 2012; Minot *et al.*, 2013). Further studies at multiple time points are needed to determine the dynamics of *P. acnes* phage populations in the skin microbiota. In addition, we found that while many individuals harbor unique phage strains (Figure 2), some individuals share the same phage strains in the skin microbiota, suggesting that a pool

of common phages exists among the human population. Lastly, we isolated identical phage strains from siblings (Supplementary Table S2), suggesting possible transmission of skin viruses between related individuals.

Pseudolysogeny occurs abundantly in natural environments and is an important part of bacteria–phage interactions (Ripp and Miller, 1997). It has been suggested to influence phage survival (Ripp and Miller, 1997), phage-dependent bacterial mortality (Ripp and Miller, 1998) and the virulence of certain bacterial strains (Sakaguchi *et al*., 2005). Although it has been observed that *P. acnes* phages can adopt pseudolysogeny (Lood and Collin, 2011), for the first time we revealed that this behavior is largely dependent on the lineages of the host strains and is not a property of the phages. All tested *P. acnes* phages more frequently become pseudolysogenic when the host *P. acnes* strains are from clade IA-1, compared with other lineages (Figure 5). We further determined several pseudolysogenic characteristics of *P. acnes* phage infection (Supplementary Figure S6). These include the ability of the phage genome to exist as a circular episome in the host cells, superinfection immunity of the host bacteria gained from prior phage infection and phage DNA persistence in a subpopulation of the host cells following infection. While the term 'pseudolysogeny' is used here and in the previous studies (Lood and Collin, 2011; Marinelli *et al*., 2012) to describe this behavior of *P. acnes* phages, further mechanistic definitions are needed to determine whether this infection state represents phages undergoing extrachromosomal lysogeny, inefficient lysis or other potential mechanisms.

We found that although *P. acnes* phages can lyse *P. acnes* strains from most lineages, certain strains from clades IB-3, II and III are resistant (Figure 4a). This bacterial lineage-dependent resistance to phages suggests that antiviral mechanisms may have a role in defining the *P. acnes* strain populations in the presence of phages. Based on the genomes of *Propionibacterium* strains, we identified two antiviral mechanisms that may explain phage resistance in different *P. acnes* lineages, restriction modification and CRISPRs. Both of these systems act against phage DNA that has successfully entered the host. It is possible that these *P. acnes* lineages employ additional mechanisms that act at other stages of infection, such as receptor binding or DNA injection, which are commonly responsible for phage resistance in mycobacteria (Jacobs-Sera *et al*., 2012).

We observed that phage DNA recognition by *P. acnes* CRISPRs often did not protect the tested bacterial strains from phage lysis (Figure 6). This result contrasts with a previous study reporting a correlation between matching spacers and phage resistance (Marinelli *et al*., 2012). While this earlier study similarly reported 1 unexplained lytic outcome out of 20 tested interactions, where 2 matched spacers failed to confer phage resistance, in our study we observed a much higher proportion of lytic outcomes, 27 out of 37 non-redundant interactions. The contrasting conclusions from the two studies may be attributed to the differences in the bacterial host strains and phage strains tested. To demonstrate the robustness of our results, we included redundant bacterial host strains and phage strains in our experiments. Both RT2 and RT6 strains, which belong to two different lineages of *P. acnes* that encode CRISPR/Cas systems, and several members of the same phage strains were tested in the analysis and yielded consistent results.

Several possible reasons may explain the low correlation between spacer matching and phage resistance. The CRISPR/Cas system in *P. acnes* may be used for purposes other than defense against foreign DNA, as suggested for the Type I-E CRISPR/Cas system of *Escherichia coli* (Touchon and Rocha, 2010; Touchon *et al*., 2011). Alternatively, other molecular regulations employed by bacterial host or phage may be important for CRISPR-mediated resistance and may contribute to the infection outcome. Such regulation mechanisms potentially include transcriptional and/or translational regulation of CRISPR RNA and Cas genes and anti-CRISPR genes encoded by phages (Bondy-Denomy *et al*., 2013; Pawluk *et al*., 2014). Interactions between *P. acnes* and phages may also depend on additional mechanisms involved in phage binding, entry, replication or release.

To our knowledge, this study is the first large-scale investigation of bacterium–phage interaction in the human skin microbiota using both genetically defined bacterial strains and phage strains. Our study suggests that the prey–predator relationship between *P. acnes* and its phages may have a role in modulating the composition of the skin microbiota. The finding that the infection outcome of *P. acnes* phages varies depending on *P. acnes* lineages suggests that the presence of phages may alter the *P. acnes* population structure by targeting particular subsets of strains. Based on evidence of different propensities among *P. acnes* lineages to be lysed, to resist or to sustain a pseudolysogenic infection, it is plausible that in the presence of phages various *P. acnes* strains may populate at different rates in the community. On the other hand, the *Propionibacterium* strain composition may determine the abundance and dynamics of the phages in the community. Phage bursts may occur in specific skin community structures. Future studies utilizing recently developed techniques, such as microfluidic digital PCR (Tadmor *et al*., 2011), phageFISH (Allers *et al*., 2013) and viral tagging (Deng *et al*., 2012), will provide further insight on the diversity and dynamics of the phage populations in the human microbiome and the prey–predator relationship between the bacterial hosts and phages.

P. acnes phages may hold therapeutic potential for modifying the skin microbiota based on the individual bacterial community structure to restore a healthy microbial composition. We previously demonstrated that *P. acnes* strain population structure differed

significantly between acne patients and healthy individuals (Fitz-Gibbon *et al.*, 2013). Certain strains were highly associated with the disease, while some strains were highly associated with healthy skin. The interactions between genetically defined *P. acnes* strains and phages that we characterized in this study may be able to guide the selection of phage strains to inhibit specific *P. acnes* population types associated with acne or other skin diseases. For example, phage therapy may not be effective if the host microbiota contains phage-resistant pathogenic strains. Potential phage-based therapy should consider the microbiome structure of the individual to be effective and personalized. The host strain specificity of *P. acnes* phages may prove advantageous for the selective suppression of pathogenic *P. acnes* strains and preservation of commensal or beneficial strains in the skin microbial community.

Conflict of Interest

The Regents of the University of California is the owner of a patent application based on some of the results presented, which names ST and HL as inventors.

Acknowledgements

This research was funded by NIH grants R01GM099530 and UH2AR057503 from NIGMS and NIAMS. JL was supported by the Microbial Pathogenesis Training Grant T32AI07323. Phage genomes were sequenced at the UCLA Genotyping and Sequencing Core. We thank Dr Emma Barnard and Emily Curd for providing assistance in preparation of MiSeq sequencing libraries, and Dr Baochen Shi for help with metagenomic data analysis.

Author contributions

JL and ST analyzed the data. RY, QZ, SN, NJB, TL and LN performed the experiments. ML helped perform the experiments. MCE collected the clinical samples. NC maintained the IRB protocol and collected the clinical samples. HL conceived and directed the project and analyzed the data. JL and HL wrote the paper.

References

Allers E, Moraru C, Duhaime MB, Beneze E, Solonenko N, Barrero-Canosa J *et al.* (2013). Single-cell and population level viral infection dynamics revealed by phage-FISH, a method to visualize intracellular and free viruses. *Environ Microbiol* **15**: 2306–2318.

Bergler-Czop B, Brzezińska-Wcisło L. (2013). Dermatological problems of the puberty. *Postepy Dermatol Alergol* **30**: 178–187.

Blaser MJ, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Estrada I *et al.* (2013). Distinct cutaneous bacterial assemblages in a sampling of South American Amerindians and US residents. *ISME J* **7**: 85–95.

Bojar R a, Holland KT. (2004). Acne and *Propionibacterium acnes*. *Clin Dermatol* **22**: 375–379.

Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**: 429–432.

Brzin B. (1964). Studies on the Corynebacterium acnes. *Acta Pathol Microbiol Scand* **60**: 599–608.

Butler-Wu SM, Sengupta DJ, Kittichotirat W, Matsen FA, Bumgarner RE. (2011). Genome sequence of a novel species, Propionibacterium humerusii. *J Bacteriol* **193**: 3678.

Chevreux B, Wetter T, Suhai S (1999). Genome sequence assembly using trace signals and additional sequence information. In: Wingender E (ed), *Computer science and biology. Proceedings of the German Conference on Bioinformatics*. GCB, Hannover, Germany, pp 45–56.

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. (2009). Bacterial community variation in human body habitats across space and time. *Science* **326**: 1694–1697.

Cummins CS. (1976). Identification of *Propionibacterium acnes* and related organisms by precipitin tests with trichloroacetic acid extracts. *J Clin Microbiol* **2**: 104–110.

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**: 4636–4641.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.

Deng L, Gregory A, Yilmaz S, Poulos BT, Hugenholtz P, Sullivan MB. (2012). Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *MBio* **3**: e00516–12.

Dorscht J, Klumpp J, Bielmann R, Schmelcher M, Born Y, Zimmer M *et al.* (2009). Comparative genome analysis of Listeria bacteriophages reveals extensive mosaicism, programmed translational frameshifting, and a novel prophage insertion site. *J Bacteriol* **191**: 7206–7215.

Duhaime MB, Deng L, Poulos BT, Sullivan MB. (2012). Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environ Microbiol* **14**: 2526–2537.

Duhaime MB, Sullivan MB. (2012). Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology* **434**: 181–186.

Egea R, Casillas S, Barbadilla A. (2008). Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res* **36**: W157–W162.

Farrar MD, Howson KM, Bojar RA, West D, Towler JC, Parry J *et al.* (2007). Genome sequence and analysis of a *Propionibacterium acnes* bacteriophage. *J Bacteriol* **189**: 4161–4167.

Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA *et al.* (2013). Topographic diversity of fungal and bacterial communities in human skin. *Nature* **498**: 367–370.

Fitz-Gibbon S, Tomida S, Chiu B-H, Nguyen L, Du C, Liu M *et al.* (2013). *Propionibacterium acnes* strain populations in the human skin microbiome associated with acne. *J Invest Dermatol* **133**: 2152–2160.

Foulongne V, Sauvage V, Hebert C, Dereure O, Cheval J, Gouilh MA *et al.* (2012). Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. *PLoS One* **7**: e38499.
Gao Z, Perez-Perez GI, Chen Y, Blaser MJ. (2010). Quantitation of major human cutaneous bacterial and fungal populations. *J Clin Microbiol* **48**: 3575–3581.
Gao Z, Tseng C, Pei Z, Blaser MJ. (2007). Molecular analysis of human forearm superficial skin bacterial biota. *Proc Natl Acad Sci USA* **104**: 2927–2932.
Gordon D, Abajian C, Green P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res* **8**: 195–202.
Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC *et al.* (2009). Topographical and temporal diversity of the human skin microbiome. *Science* **324**: 1190–1192.
Horvath P, Barrangou R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**: 167–170.
Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle MM *et al.* (2012). On the nature of mycobacteriophage diversity and host preference. *Virology* **434**: 187–201.
Jong EC, Ko HL, Pulverer G. (1975). Studies on bacteriophages of *Propionibacterium acnes*. *Med Microbiol Immunol* **161**: 263–271.
Katoh K, Misawa K, Kuma K, Miyata T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066.
Kong HH, Oh J, Deming C, Conlan S, Grice EA, Beatson MA *et al.* (2012). Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res* **22**: 850–859.
Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C *et al.* (2004). Versatile and open software for comparing large genomes. *Genome Biol* **5**: R12.
Leyden JJ. (2001). The evolving role of *Propionibacterium acnes* in acne. *Semin Cutan Med Surg* **20**: 139–143.
Li H. (2010). Metagenomic study of the human skin microbiome associated with acne. *J Invest Dermatol* **133**: 2152–2160.
Liu J, Cheng A, Bangayan NJ, Barnard E, Curd E, Craft N *et al.* (2014). Draft genome sequences of *Propionibacterium acnes* type strain ATCC6919 and antibiotic-resistant strain HL411PA1. *Genome Announc* **2**: e00740–14.
Lomholt HB, Kilian M. (2010). Population genetic analysis of Propionibacterium acnes identifies a subpopulation and epidemic clones associated with acne. *PLoS One* **5**: e12277.
Lood R, Collin M. (2011). Characterization and genome sequencing of two Propionibacterium acnes phages displaying pseudolysogeny. *BMC Genomics* **12**: 198.
Lukashin AV, Borodovsky M. (1998). GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* **26**: 1107–1115.
Ma Y, Madupu R, Karaoz U, Nossa CW, Yang L, Yooseph S *et al.* (2014). Human papillomavirus community in healthy persons, defined by metagenomics analysis of human microbiome project shotgun sequencing data sets. *J Virol* **88**: 4786–4797.
Manica A, Zebec Z, Steinkellner J, Schleper C. (2013). Unexpectedly broad target recognition of the CRISPR-mediated virus defence system in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res* **41**: 10509–10517.
Marinelli LJ, Fitz-gibbon S, Hayes C, Bowman C, Inkeles M, Loncaric A *et al.* (2012). *Propionibacterium acnes* bacteriophages display limited genetic diversity and broad killing activity against bacterial skin isolates. *MBio* **3**: e00279–12.
Marples RR, McGinley KJ, Mills OH. (1973). Microbiology of comedones in acne vulgaris. *J Invest Dermatol* **60**: 80–83.
McDowell A, Barnard E, Nagy I, Gao A, Tomida S, Li H *et al.* (2012). An expanded multilocus sequence typing scheme for propionibacterium acnes: investigation of 'pathogenic', 'commensal' and antibiotic resistant strains. *PLoS One* **7**: e41480.
Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. (2013). Rapid evolution of the human gut virome. *Proc Natl Acad Sci USA* **110**: 12450–12455.
Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. (2013). Expanding the marine virosphere using metagenomics. *PLoS Genet* **9**: e1003987.
Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733–740.
Nakatsuji T, Chiang H-I, Jiang SB, Nagarajan H, Zengler K, Gallo RL. (2013). The microbiome extends to subepidermal compartments of normal skin. *Nat Commun* **4**: 1431.
Nolan JM, Petrov V, Bertrand C, Krisch HM, Karam JD. (2006). Genetic diversity among five T4-like bacteriophages. *Virol J* **3**: 30.
Ördögh L, Hunyadkürti J. (2013). Complete genome sequence of Propionibacterium avidum strain 44067, isolated from a human skin abscess. *Genome Announc* **1**: 5–6.
Paulino LC, Tseng C-H, Blaser MJ. (2008). Analysis of Malassezia microbiota in healthy superficial human skin and in psoriatic lesions by multiplex real-time PCR. *FEMS Yeast Res* **8**: 460–471.
Pawluk A, Bondy-Denomy J, Cheung VHW, Maxwell KL, Davidson AR. (2014). A new group of phage Anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *MBio* **5**: e00896–14.
Pride DT, Salzman J, Haynes M, Rohwer F, Davis-Long C, White RA *et al.* (2012). Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *ISME J* **6**: 915–926.
Puhvel SM, Amirian DA. (1979). Bacterial flora of comedones. *Br J Dermatol* **101**: 543–548.
Ripp S, Miller R V. (1998). Dynamics of the pseudolysogenic response in slowly growing cells of *Pseudomonas aeruginosa*. *Microbiology* **144**: 2225–2232.
Ripp S, Miller R V. (1997). The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. *Microbiology* **143**: 2065–2070.
Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pasić L, Thingstad TF, Rohwer F *et al.* (2009). Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7**: 828–836.
Rohwer F. (2003). Global phage diversity. *Cell* **113**: 141.
Rohwer F, Thurber RV. (2009). Viruses manipulate the marine environment. *Nature* **459**: 207–212.
Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA *et al.* (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945.
Sakaguchi Y, Hayashi T, Kurokawa K, Nakayama K, Oshima K, Fujinaga Y *et al.* (2005). The genome

sequence of Clostridium botulinum type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. *Proc Natl Acad Sci USA* **102**: 17472–17477.

Semenova E, Jore MM, Datsenko K a, Semenova A, Westra ER, Wanner B *et al.* (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci USA* **108**: 10098–10103.

Semenova E, Nagornykh M, Pyatnitskiy M, Artamonova II, Severinov K. (2009). Analysis of CRISPR system function in plant pathogen Xanthomonas oryzae. *FEMS Microbiol Lett* **296**: 110–116.

Sepúlveda-Robles O, Kameyama L, Guarneros G. (2012). High diversity and novel species of *Pseudomonas aeruginosa* bacteriophages. *Appl Environ Microbiol* **78**: 4510–4515.

Solonenko SA, Ignacio-Espinoza JC, Alberti A, Cruaud C, Hallam S, Konstantinidis K *et al.* (2013). Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genomics* **14**: 320.

Suttle CA, Chan AM, Cottrell MT. (1990). Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* **347**: 467–469.

Tadmor AD, Ottesen EA, Leadbetter JR, Phillips R. (2011). Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science* **333**: 58–62.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

The Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 207–214.

Tomida S, Nguyen L, Chiu B-H, Liu J, Sodergren E, Weinstock GM *et al.* (2013). Pan-genome and comparative genome analyses of propionibacterium acnes reveal its genomic diversity in the healthy and diseased human skin microbiome. *MBio* **4**: e00003–e00013.

Touchon M, Charpentier S, Clermont O, Rocha EPC, Denamur E, Branger C. (2011). CRISPR distribution within the Escherichia coli species is not suggestive of immunity-associated diversifying selection. *J Bacteriol* **193**: 2460–2467.

Touchon M, Rocha EPC. (2010). The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. *PLoS One* **5**: e11126.

Ursell LK, Clemente JC, Rideout JR, Gevers D, Caporaso JG, Knight R. (2012). The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites. *J Allergy Clin Immunol* **129**: 1204–1208.

Waterbury JB, Valois FW. (1993). Resistance to co-occurring phages enables marine synechococcus communities to coexist with cyanophages abundant in seawater. *Appl Envir Microbiol* **59**: 3393–3399.

Webster GF, Cummins CS. (1978). Use of bacteriophage typing to distinguish propionibacterium acnes types I and II. *J Clin Microbiol* **7**: 84–90.

Westra ER, van Erp PBG, Künne T, Wong SP, Staals RHJ, Seegers CLC *et al.* (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by cascade and Cas3. *Mol Cell* **46**: 595–605.

White GM. (1998). Recent findings in the epidemiologic evidence, classification, and subtypes of acne vulgaris. *J Am Acad Dermatol* **39**: S34–S37.

Wylie KM, Mihindukulasuriya KA, Zhou Y, Sodergren E, Storch GA, Weinstock GM. (2014). Metagenomic analysis of double-stranded DNA viruses in healthy adults. *BMC Biol* **12**: 71.

Zierdt CH, Webster C, Rude WS. (1968). Study of the anaerobic corynebacteria. *Int J Syst Bacteriol* **18**: 33–47.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)

Chapter 6

Metagenomic analysis of *Propionibacterium acnes* bacteriophage populations in the human skin microbiome

Abstract

Bacteriophages that infect *Propionibacterium acnes* may play a role in regulating the population composition of this dominant skin commensal bacterium, which has been implicated in acne vulgaris and many other diseases. While previous studies have characterized the diversity and host range of isolated phages, the dynamics and structure of *P. acnes* phage populations remain to be fully characterized. In this study, metagenomic and culture-based analyses of phage populations in healthy individuals and acne patients reveal that while phages are generally prevalent at varying abundance across subjects, phage populations are highly personal, and limited genetic change occurs in phage strains over time. The skin metagenomes of two related individuals revealed more genetically similar phage populations than those of unrelated individuals as well as a potential transmission of a phage between these subjects. These results shed light on the inter-individual and temporal characteristics of *P. acnes* phage populations, which may influence the effectiveness of phage therapy.

Introduction

Propionibacterium acnes is a dominant commensal skin bacterium that is implicated in many postoperative infections as well as inflammatory diseases (60), such as the common skin disease acne vulgaris. Previous studies have identified *P. acnes* strains associated with healthy or acne-affected skin (61), as well as putative virulence factors in acne-associated strains that may confer pathogenic properties (52, 62). However, the factors that influence the proportions of health-associated, acne-associated, and commensal strains of *P. acnes* in an individual's skin microbiota are still largely unknown.

Bacteriophages of *P. acnes* have attracted increasing interest for their potential roles in regulating *P. acnes* populations (63) and as a potential therapeutic against pathogenic *P. acnes*

(64). Recent studies have laid the groundwork for understanding these roles by characterizing the natural diversity of phages as well as their interaction with strains of *P. acnes.* Culture-dependent analysis of phage diversity showed high genetic similarity among isolated phage strains (65), and we additionally found that while most phage strains are unique to individuals, some phage strains can also be found in different individuals (63). Phages can broadly infect many strains of *P. acnes* (65), and the lytic activity of these phages depends on the lineage of *P. acnes* host (63).

These initial observations of phage diversity and host interaction now motivate an investigation of the distribution and composition of phage populations, which may affect the corresponding population structure of their *P. acnes* hosts. In this study, we complement our culture-based approaches with metagenomic analyses to determine the abundance and structure of phage populations across the human population and within individuals.

We also consider the transmission of phage strains between humans, which has intuitive importance as a process that affects the assembly of phage populations in new hosts as well as the ecology of established phage populations. In a previous study, we found that siblings share the same or highly similar phage strains in pilosebaceous units (63). Here, we report metagenomic and culture-based evidence suggesting the transmission of a phage strain between related individuals.

Methods

Subjects

Subjects with acne or healthy skin were recruited from various clinics in Southern California, USA, including private practice, managed care, and public hospital settings, as well as outside of dermatology clinics, to best represent the diversity of populations and history of

medical care. The diagnosis of acne was made by board-certified dermatologists. The presence of acne was graded on a scale of 0 to 5 according to the Global Acne Severity Scale (66). Grades were recorded for both the face and the nose separately, where zero represents healthy skin and 5 represents the most severe inflammatory cystic acne. The presence of scarring was also noted. Subjects with healthy skin were determined by board-certified dermatologists and were defined as people who had no acneiform lesions on the face, chest, and back. They were excluded if they had other skin conditions that the investigators felt would affect sampling or the microbial population on the skin. Additionally, healthy subjects were excluded if they had been previously treated for acne. The subjects responded to a written questionnaire, administered by a physician or a well-trained study coordinator who went over each question with the subjects. All subjects provided written informed consent. All protocols and consent forms were approved by both the UCLA and Los Angeles Biomedical Research Institute IRBs. The study was conducted according to the Helsinki Guidelines.

Sample collection and genomic DNA extraction

Sample collection and processing were performed as previously described in Fitz-Gibbon et al. (61). Briefly, skin follicle samples were taken from the nose of the subjects using Bioré Deep Cleansing Pore Strip (Kao Brands Company, Cincinnati, OH) following the manufacturer's instructions. Clean gloves were used for each sampling. To avoid sample cross-contamination, strict sterile working conditions were enforced, including appropriate lab-wear and routine sterilization of the working surfaces and equipment. After being removed from the nose, the strip was immediately placed into a 50 ml sterile tube and kept on ice or at 4°C. Samples were processed within 24 hours of collection. Individual follicular plugs were manually picked from the adhesive strip with sterile forceps and placed into 2 ml sterile microcentrifuge tubes filled

with ATL buffer (Qiagen) and glass beads (0.1 mm diameter) (BioSpec Products Inc., Bartlesville, OK). Cells were lysed using a beadbeater for 3 minutes at 4,800 rpm at room temperature, then centrifuged at 14,000 rpm for 5 minutes. The supernatant was retrieved and used for genomic DNA (gDNA) extraction using QIAamp DNA micro kit (Qiagen). The manufacturer protocol for extracting genomic DNA from chewing gum was used. DNA extraction kit reagents were routinely tested for contaminants using negative control extractions. Concentration of gDNA was determined using Qubit Fluorometric Quantitation (Invitrogen). Concentrations from negative control experiments were below the detectable range.

Library construction and metagenomic shotgun sequencing

Genomic sequencing libraries were prepared using NexteraXT kit (Illumina). Briefly, 1 ng of extracted metagenomic DNA was tagmented using transposase technology. 5' and 3' NexteraXT dual-indices were added to uniquely barcode each library. Indexed libraries were amplified using the limited 12-cycle PCR program as instructed by the manufacturer's guidelines. Libraries were purified with Agencourt AMPure XP magnetic beads (Beckman Coulter). Library quality and average fragment length were assessed using Bioanalyzer (Agilent Technologies). Libraries were quantified with KAPA library quantification kit (KAPA Biosystems), according to the manufacturer's instructions. Finally, libraries were randomly pooled together and sequenced using MiSeq and/or HiSeq platforms (Illumina) with paired-end reads of 251 bp or 101 bp, respectively. Following sequencing, reads were de-multiplexed.

Data cleaning and sequencing depth analysis

Data cleaning of the metagenomic shotgun sequences was performed following the protocol used in the HMP (67). Briefly, low quality reads and human DNA reads were filtered

out first, resulting in 136 Gbp in total for 140 samples (average 0.97 Gbp per sample). The remaining reads were mapped against reference genomes. The average number of mappable reads per sample was 4.86×10^8 (1.12×10^7 – 4.14×10^9 reads), which were used in the downstream analyses.

Phage isolation

P. acnes phages were isolated and sequenced as previously described (63). Briefly, skin follicle samples were spread onto blood agar (Teknova Brucella Agar Plate with Hemin and Vitamin K, Teknova, Hollister, CA, USA) and incubated at 37°C for 5 – 6 days. Material from observed plaques was suspended in 50 µL SM buffer (0.1M sodium chloride, 8mM magnesium sulfate heptahydrate, 1M Tris-HCl, pH 7.5, 2% gelatin and 1mM calcium chloride) and spread onto A-media plates (12 g/L pancreatic digest of casein, 12 g/L yeast extract, 22.2mM D-glucose, 29.4 mMg/L potassium phosphate mono-basic, 8mM magnesium sulfate heptahydrate and 20 g/L agar) with top agar containing *P. acnes* strain ATCC6919. After incubation at 37 °C for 2 days, single plaques were selected and propagated using the same host strain, medium and incubation conditions. Suspensions of phage isolates were prepared by eluting plates with 8 mL SM Buffer and filtering with a 0.22-µm PES filter (Millipore, Billerica, MA, USA). Phage DNA extraction was performed using Lambda Mini Kit (Qiagen, Valencia, CA, USA) with some modifications to the original protocol (63).

Phage genome comparison

The core genomic regions among phage isolate genomes or phage population consensus sequences were calculated as previously described (63). Briefly, Nucmer (68) was used to determine sets of genomic regions within a reference genome (here the PA6 genome) that show

homology to each of the other genomes. The core regions were then computed as the overlapping regions between all sets. Single nucleotide polymorphisms (SNPs) output by Nucmer were used to construct a phylogenetic tree in MEGA 5 using the Neighbor-Joining method on p-distances (53). Bootstrapping was performed using 2,000 replicates.

Phage population analysis

P. acnes phage reads were obtained by using Bowtie2 (69) to map total microbial shotgun sequencing reads for each sample to a reference 'pan-genome' of non-redundant phage genomic regions. The pan-genome was constructed from 62 publically available phage genomes as of November 5, 2015, using methods previously described (52). The phage reads for each sample were *de-novo* assembled into a metagenomic phage population assembly using Mira 3.2.1 (45). Shotgun metagenomic data from Human Microbiome Project (HMP) samples SRS017849 and SRS017851 were analyzed according to the above methods, except that the reference genome of phage PA6 was used as the mapping reference. Access to the phenotype information for HMP data (phs000228.v3.p1) was obtained from dbGaP. Phage relative abundance in each sample was calculated by normalizing the total bases mappable to the phage pan-genome by the average phage genome length of 29,434 nt previously reported in Liu et al (63). This normalized base count was then divided by the sum of the normalized base counts for all reference genomes in the sample that had >1% of their genome length covered by mapped reads.

Population consensus sequences were manually finished in Consed 26.0 (70). Nucleotide variations within the core regions shared between each consensus sequence were determined as

previously described (63).SNPs were identified in each assembly as positions covered by ≥ 30 reads (with Phred quality ≥ 30) with ≥ 10% of these reads differing from the consensus.

To identify allele variations between pairs of samples, a Perl script was used to calculate allele frequencies at SNP sites within the ACE-formatted, metagenomic assembly of phage reads in one sample. Next, the population consensus sequence of the sample was aligned with the consensus of the second sample using MUSCLE (54). The allele frequencies at SNP sites of the first sample were then compared to the allele frequencies at sites of the second sample that aligned with the SNP sites of the first sample.

Results

Prevalence of phage populations

We investigated the inter-individual differences and intra-individual dynamics of *P. acnes* phage populations by performing metagenomic shotgun sequencing of 140 samples from 103 individuals. This sample set included 31 longitudinal sample pairs collected from the same individuals at two time points separated by 10 –54 weeks, 24 samples collected from 20 related individuals, and 10 samples collected from different sites of 4 individuals.

P. acnes phages were found to be prevalent, but in varying abundance, within our study cohort. We quantified the relative abundance of phages in each sample as the number of bases that could be mapped to a pan-genome constructed from 62 sequenced phage genomes normalized by the average phage genome length. *P. acnes* phages were detected in 73 of the 140 samples (52%), with a cutoff of 1× coverage of the average phage genome length (29,434 nt, see Methods). Phage abundances were more similar between different sites of the same individual

compared with the nose sites sampled from different individuals (Figure 6-1), though this is based on a limit number of samples.

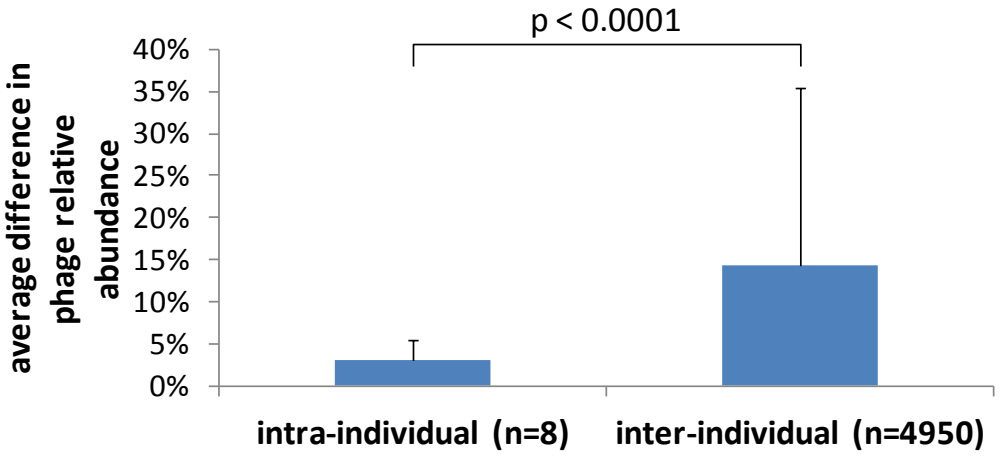


Figure 6-1. The abundance of *P. acnes* phages at different sites within the same subject (chin, nose, forehead) was more similar than between the nose sites of different individuals. Averages are of all pairwise differences in phage abundance between samples in each group.

Diversity between and within phage populations

To compare the inter- and intra-individual genetic similarity between phage populations, we performed *de novo* assembly of the phage reads in each of 33 samples that had ≥11× coverage of the average phage genome length (Table 6-1). These samples were collected from a total of 20 individuals and include samples from 5 individuals collected at two time points, samples from two related individuals, and samples from two facial sites of one individual. These assemblies yielded 34 nearly full-length consensus sequences of the underlying phage populations, one from each sample except HL213M, where two distinct consensus genome sequences were obtained. We compared the resulting 34 population consensuses by calculating the pairwise nucleotide differences in the core regions shared among these sequences.

Table 6-1. Assembly statistics for metagenomic samples ≥ 11× coverage on the average phage genome length

Sample	Longitudinal pair	Sequencing Coverage	Assembly coverage	SNPs	Strains in population
HL074N	1	99.9	97.6	476	>1
HL074M	1	36.4	33.5	553	>1
HL073M	2	61.7	61.3	11	1
HL073N	2	619.2	614.1	132	>1
HL081N	3	218.3	216.6	31	1
HL081M	3	17.9	17.0	10	1
HL009M	4	168	165.1	987	>1
HL009N	4	115.8	113.0	664	>1
HL041M	5	173.4	165.4	6	1
HL041N	5	173.1	171.1	0	1
HL066M		50.2	49.3	5	1
HL313M-FOREHEAD		41.3	40.8	0	1
HL313M-CHIN		39.8	37.5	0	1
HL085N		448.5	444.0	6	1
HL213M		440.6	364.9 56.7	0 0	>1
HL003N		413.1	410.4	0	1
HL214M		136.8	136.3	0	1
HL088M		137.4	134.1	1	1
HL077M		116.7	116.0	0	1
HL069N		87.7	87.0	0	1
HL070N		48.7	47.3	1,237	>1
HL091N		45.3	44.8	5	1
HL006N		36.6	36.4	0	1
HL145N		36.3	35.8	115	>1
HL315M		36.4	28.6	101	>1
HL212M		22.5	21.2	29	1
HL092N		21.3	20.9	0	1
HL075N		19.3	19.1	0	1
HL102M		16.0	15.8	0	1
HL080N		14.9	14.8	0	1
HL089M		13.0	12.6	0	1
HL068N		12.4	11.2	0	1
HL079N		11.9	10.4	24	1

The genetic similarity between phage populations was highest within an individual as well as between related individuals. Most of the 34 population consensuses are genetically distinct from each other (Figure 6-2A) by an average of 832 core nucleotide differences (range 0 – 1201). However, fewer core nucleotide differences were observed between the consensus sequences of samples from the two related individuals (24 differences, Figure 6-2B) as well as between the two different facial sites of one individual (no differences). Longitudinal sample

pairs from 5 individuals also had significantly fewer core nucleotide differences (5 – 475, average 179) than non-longitudinal sample pairs from unrelated individuals (34 – 1133, average 833, p=0.002; Figure 6-2B). This suggests that different facial sites of the same individual harbor the same *P. acnes* phage populations and that the phage populations are temporally stable. The observation that related individuals are colonized by genetically similar phage strains also suggests potential transmissions of the phage, or the bacterial host carrying the phage, between individuals. We verified the similarity of phage populations within an individual in an independent cohort by analyzing the metagenomic samples from the left and right retroauricular sites from one subject sampled by the HMP. The consensus sequences of phage populations at these two sites were highly similar, differing by only 210 nucleotides. Together, these results suggest that *P. acnes* phage populations within an individual exhibit greater genetic similarities over temporal and spatial dimensions than phage populations between individuals.

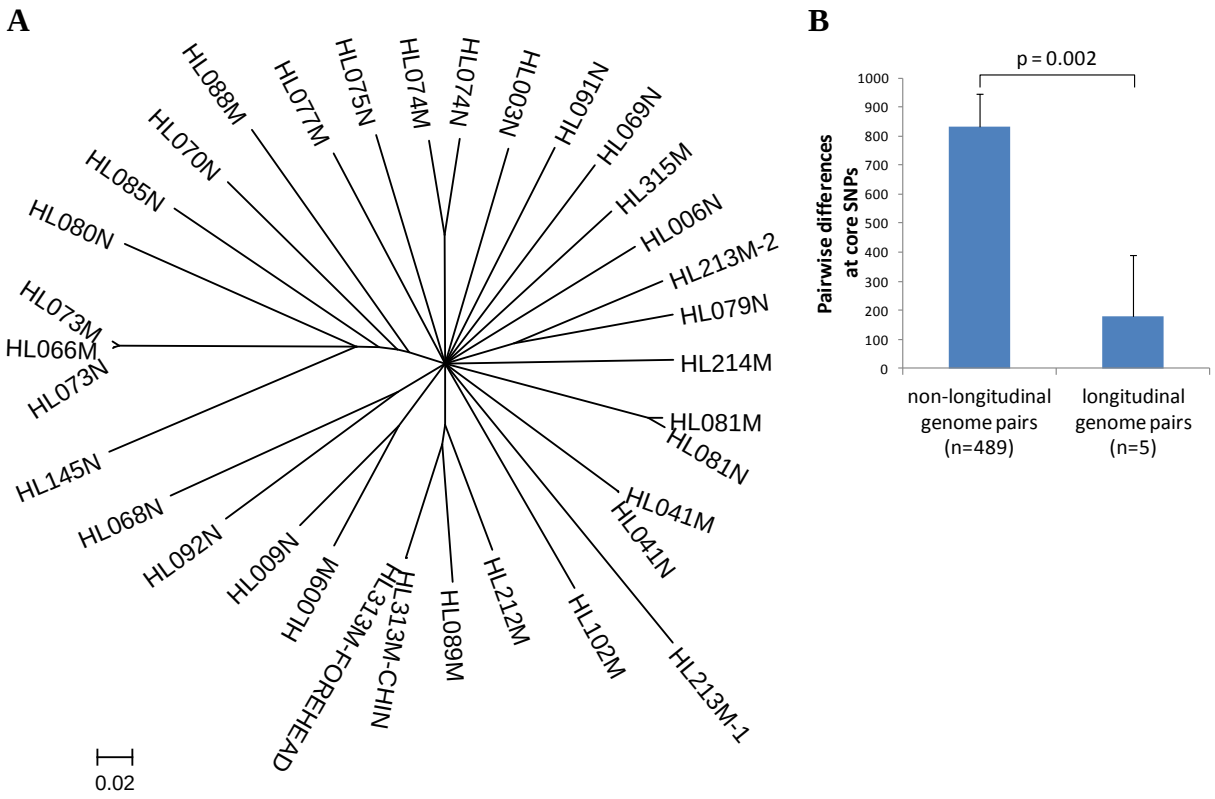


Figure 6-2. Phage populations are more similar within individuals than between individuals. (A) Phylogenetic tree constructed from nucleotide differences at 3,011 core SNP sites between the 34 phage population consensus sequences. (B) Average nucleotide differences at core SNPs within the population consensus sequences from longitudinal sample pairs from 5 individuals compared with samples from 21 other unrelated individuals. In both analyses, a single population consensus sequence was chosen to represent the phage population from a subject sampled at two different sites as well as a sample from which two distinct population consensus sequences were assembled.

We observed different degrees of genetic variation within individual phage populations. We identified 0 – 1,237 SNP positions in the core regions of each of the 33 metagenomic phage assemblies (Table 6-1). While some samples (4 of 33) contained more than 400 detected SNPs, most samples (29 of 33) contained fewer than 132 detected SNPs. This suggests that most

individuals are dominated by a single phage strain, with few individuals carrying more than one strain, as observed in our previous culture-based study (63).

Slow genetic dynamics in phage strains over time

We investigated the short-term and long-term genetic changes of phage strains on the skin. We compared the genome sequences from a collection of 31 phages isolated from 7 individuals at time points separated by intervals as short as 2 days to as long as 2.5 years. This collection consists of 7 phages isolated at different time points from our previous study (63) as well as 24 phages isolated from healthy individuals. The phages from each individual isolated from different time points are highly similar to each other (Figure 6-3), differing by no more than 17 nucleotides. This is comparable to the ~14 nucleotides of background genetic polymorphism previously observed within a clonal strain population (63) and suggests that a phage strain can persist in the skin microbiota for long periods of time.

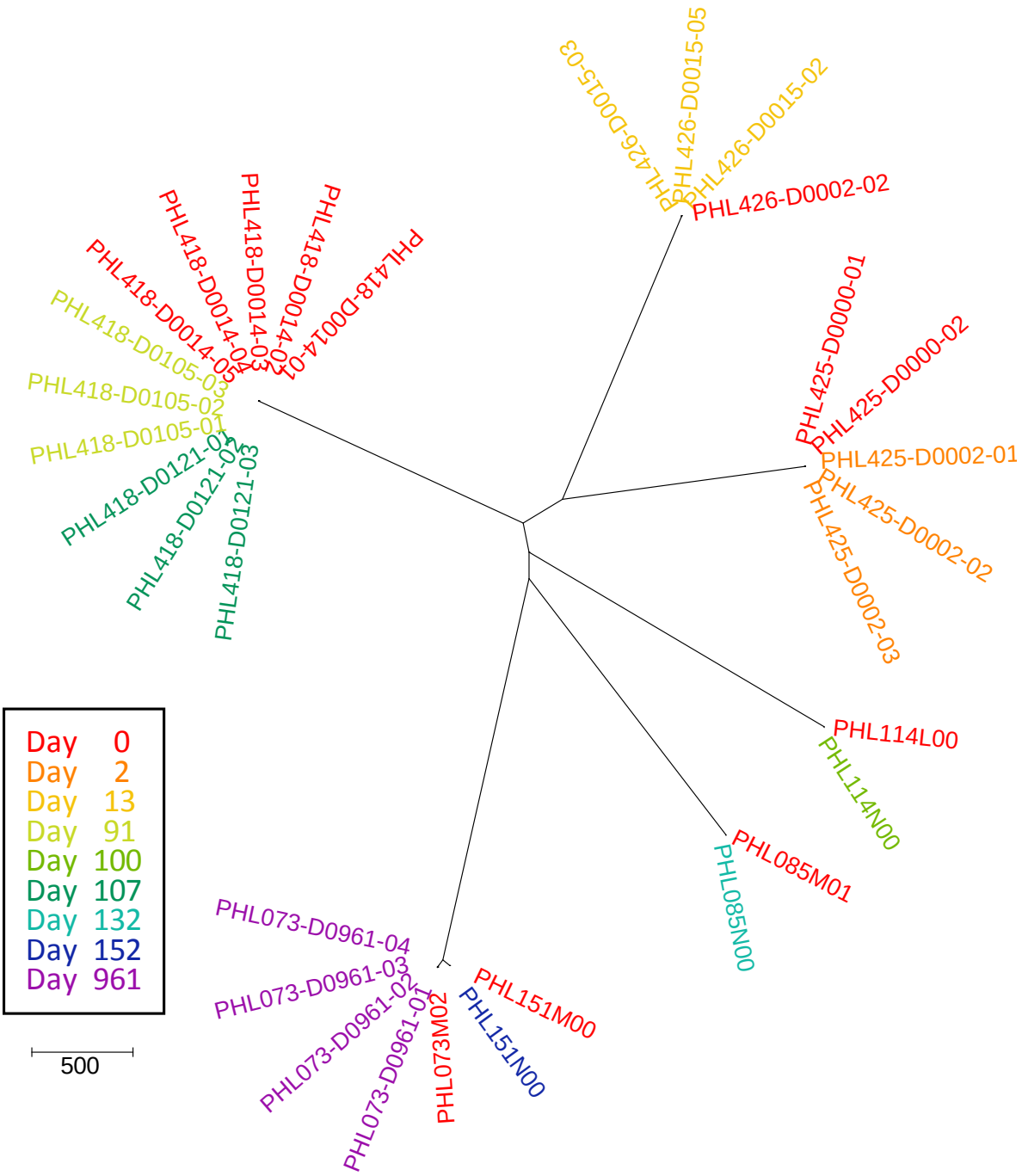


Figure 6-3. Limited genetic change in phage strains over short and long intervals. 31 phages isolated from 7 individuals at different time points were sequenced, and a phylogenetic tree was constructed from a whole –genome multiple alignment. Phages isolated from each individual form distinct clusters based on genetic similarity. The color of each phage represents the days elapsed since the first phage from the individual was isolated. Branch length represents number of nucleotide differences. Bootstrap values are based on 500 replicates.

Phage similarity between related individuals

To further investigate the similarity between the phage populations in related individuals, we used metagenomic sequencing to study the dynamics of phage populations in siblings HL010, HL066 and HL073. In my previous study (63), we found that phages isolated from HL010 and HL066 were the same strain (represented by phages in Group VIII), as their genomes differ by less than 10 nucleotides. The phage from HL073 was similar to the Group VIII strain, as 81 nucleotide differences separated the genome of PHL073M02 from the genomes of Group VIII isolates. We collected nose follicular samples from HL066 and HL073 and mapped metagenomic reads from these samples to a reference pan-genome constructed from 62 sequenced phage genomes. This yielded phage reads totaling more than 50× coverage (50.2× on HL066 and 61× on HL073) on the average phage genome. Phage relative abundance was 41% in HL066 and 85% in HL073. We also analyzed a sample collected from HL073 during a follow-up visit 10 weeks later by metagenomic shotgun sequencing. Phages made up 63% of the total microbial community based on 619× coverage in phage reads recovered from this sample.

The phage population in HL073 became more genetically similar to HL066 at the second sampling time. Few core SNP sites were detected in the assembled phage reads from each of the initial samples from these subjects, which we denote as HL066M (5 SNPs) and HL073M (11 SNPs, Table 6-1), suggesting that these individuals are colonized by single phage strains. Moreover, these phage strains are genetically similar, but distinct from each other, as the consensus sequences from these assemblies differed by 110 nucleotides (Figure 6-5). However, the phage assembly from the follow-up sampling of HL073, denoted as HL073N, contained a larger number of SNPs (132) than the initial sample from this individual, and the consensus sequence of this assembly differed by fewer nucleotides from the consensus of HL066M (10

differences) than from the consensus of HL073M (104 differences) (Figure 6-4, 6-5). These observations raise the possibility that, at the second visit, the phage population of HL073 harbored a second phage strain that is genetically similar to the one in HL066 potentially via the transmission of phages or phage-infected bacteria between these siblings.

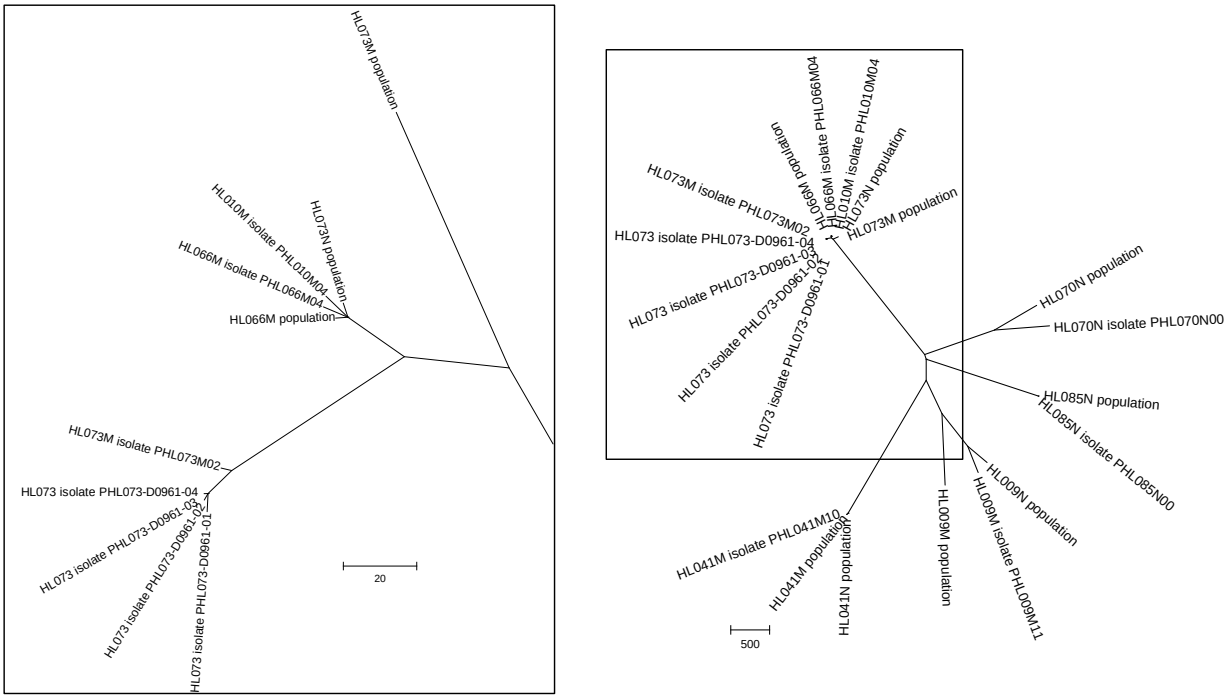


Figure 6-4. Increased genetic similarity between the phage populations of related individuals HL066 and HL073 during a later sampling of HL073. A phylogenetic tree was generated from a multiple alignment of population consensus sequences and isolate genomes from related individuals HL010, HL066, and HL073 along with the population consensus and available isolate genomes from the four other longitudinal samples. Inset shows a magnified view of the relatedness between the population consensuses of the later HL073 sample and the initial HL066 sample. Branch length represents number of nucleotide differences. Bootstrap values based on 500 replicates.

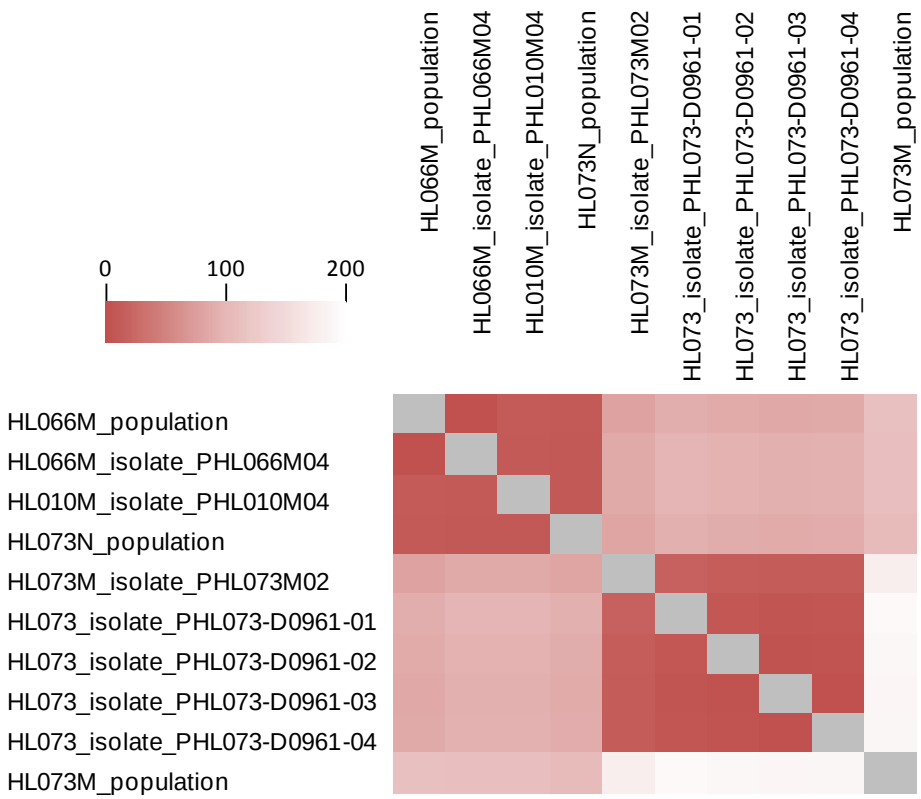


Figure 6-5. Genome-wide nucleotide differences between population consensus and isolate sequences from related individuals HL010, HL066, and HL073.

We further confirmed the presence of two strains in HL073N by comparing longitudinal allele changes in this phage population. Alignment of the population consensus sequences from HL073M and HL073N revealed 104 nucleotide changes that occurred in the population between the sampling times. 94 of these changes correspond to bi-allelic SNP sites in HL073N where the major allele is the base at the corresponding aligned position of the HL066M consensus (81% average allele frequency), while the minor allele is the base at the corresponding aligned position of the HL073 consensus (17% average allele frequency). This suggests that during the 10 weeks between the samplings the phage population of HL073, which was initially colonized by a single

strain as shown above, became dominated by another phage strain, one that was potentially transmitted from siblings.

Despite the rise in abundance of a previously undetected phage strain in HL073, lowly-abundant phage strains can still persist in the skin microbiota of this subject. PHL073M02, a phage isolated from sample HL073M, differs by 177 nucleotides from the HL073M population consensus but does not contribute to its detectable genetic variation (only 11 detected SNPs in this sample, as described above), indicating that this phage is likely present at low abundance in this population. We isolated four additional phages (PHL073-D0961-01, PHL073-D0961-02, PHL073-D0961-03, and PHL073-D0961-04) from a third follow-up sampling of HL073 2.4 years after HL073N (and 2.6 years after HL073M). The genomes of these four isolates are nearly identical (1 –7 pairwise nucleotide differences; Figure 6-4) and only slightly diverged from PHL073M02 (13 –17 nucleotide differences), suggesting that this phage strain may have maintained a consistent presence in this subject.

Discussion

Building upon earlier studies that characterized *P. acnes* phage diversity and host range (63, 65), our present study reveals several features of *P. acnes* phage populations within and between individuals that may be useful for understanding its role in disease and its potential therapeutic applications as a probiotic.

P. acnes phage populations show more intra-individual similarity than inter-individual similarity with respect to genetic relatedness and abundance. This result is consistent with the results of the previous chapter describing how phages isolated from related individuals or the same individual at different times are often the same or similar strains (63). In that chapter, the

same phage strain was also isolated from multiple unrelated individuals, whereas no two metagenomic samples from unrelated individuals in the present study consist of the same strain. However, as the number of samples containing adequate phage reads is limited in this study, further investigation is needed to determine whether these inter-individually shared isolates come from phage populations that are identical in different individuals or whether they represent transiently shared phage strains. The longitudinal similarity and often single-strain composition of phage populations, as well as the limited genetic variation we observed in phage strains over time, may suggest that phage populations, established either naturally or for therapeutic purposes, are generally stable in an individual, though further investigation is needed to determine the factors that affect phage prevalence and persistence in the human population.

Our metagenomic sequencing data suggested that phage strains may be transmitted between siblings. Such transfers of phage strains between individuals may explain why the same phage strain can be isolated from multiple unrelated individuals (as noted above). Although phage populations can be dominated by a single phage strain, the long-term persistence of one or more low-abundance strains suggests that the dynamics of potential transmission may be complex. From a therapeutic perspective, this observation also raises the possibility that probiotic phages could be transmitted from a treated patient to other individuals with as yet unknown consequences. Further investigation is needed to determine the nature of potential exchanges of phages, which may occur through transmission of phage particles or infected bacterial hosts, as well as whether such events eventually resolve into single-strain populations or remain as stable multi-strain populations.

Chapter 7

Conclusions

The effectiveness of probiotics may be improved by a better understanding of the microbial communities in which they function. In elucidating the ecological and biological aspects of potential bacterial and phage probiotics, my Ph.D. studies described in the previous chapters offer insights on the potential therapeutic use of *L. johnsonii* and bacteriophages of *P. acnes*. These findings merit further investigation in the context of these microbes and in the larger context of other probiotics.

Divergent modulation of systemic genotoxicity by *L. johnsonii* and the gut microbiota: treatment context may matter

In support of a study investigating the influence of the gut microflora on systemic genotoxicity, we observed that *L. johnsonii* strain 456 could reduce inflammation and systemic genotoxicity when inoculated into a DNA repair-deficient mouse model or wild-type mice raised under conventional housing conditions. However, treatment with 456 increased radiation-induced DNA-damage. Moreover, we observed that a restricted microflora, previously shown to reduce systemic genotoxicity, was also associated with higher levels of radiation-induced DNA damage. In light of these results, the genotoxicity-reducing effects of 456 and of specific gut microflora compositions may depend on the conditions under which they are applied or evaluated. The results of our study suggest that 456 may be beneficial as an oncoprotective supplement but potentially harmful as a radiotherapy adjuvant. A larger study cohort with more robust controls is needed to further confirm these contrasting effects as well as determine the mechanisms by which they arise.

Comparative genome analysis can provide insights into probiotic function

Comparative genomics of *L. johnsonii* strain 456 opens potential avenues for investigating its probiotic activity. Analysis of conserved *L. johnsonii* genes and core genomic

regions among sequenced *L. johnsonii* strains demonstrated the genetic distinction of 456 from other mouse isolates. Within the 456 genome, we also found genes exclusive to a subset of *L. johnsonii* isolates or to the 456 genome alone that may alter attachment or interaction of this strain with the host. As these conclusions are currently limited by the draft quality of the 456 genome and the small number of *L. johnsonii* genome sequences available for comparison, additional sequencing of 456 and other *L. johnsonii* strains isolated from other hosts will be needed to determine the host-specific adaptations and compatibility of strains from non-native hosts.

Diversity and host interactions of *P. acnes* bacteriophages: microbiome structure may also matter

By sequencing phage isolates from a cohort of healthy subjects and acne patients and analyzing metagenomic sequencing data from healthy skin metagenomes, we found that *P. acnes* phages are prevalent among the human population, and individuals that harbor phage populations are often colonized by a single *P. acnes* phage strain. We also found that phage infection could result in lysis, resistance, or pseudolysogeny in *P. acnes* hosts depending on the lineage of the bacterial strains tested, suggesting that the combination of *P. acnes* lineages and phage strains present in an individual's skin microbiome may determine the overall response to phage-based therapy. Further investigation is needed to better understand the role of phages in the pathogenesis of *P. acnes*-associated diseases, for instance through correlation of phage abundance with *P. acnes* strain composition, as well as through observing the direct ecological effect of phages on *P. acnes* strain composition, which can be studied through in-vitro infections of mock *P. acnes* populations with single or multiple phage strains.

Phage dynamics: temporal stability and transmissibility of probiotic treatment

Further analysis of the skin metagenomes from healthy subjects and acne patients confirmed our previous observations on phage diversity and revealed that phage populations within an individual are temporally stable, with limited genetic change occurring within a phage strain across time. However, in one subject, we also found metagenomic evidence of a phage strain, potentially transmitted from related individuals, that became the dominant strain in the phage population. These results suggest that while phage populations can remain genetically stable for long durations, they are also subject to drastic changes due to, for instance, changes in strain composition.

Characterizing the dynamics of phage populations remains a challenging but therapeutically relevant goal. Approximately half of the subjects from our study cohort showed evidence of harboring *P. acnes* phages based on metagenomic sequencing. Considering the limited biomass of skin samples, extensive sampling of a large number of subjects is recommended for investigating the ecology of these phage populations.

Appendix

Draft Genome Sequences of *Propionibacterium acnes* Type Strain ATCC6919 and Antibiotic-Resistant Strain HL411PA1

Jared Liu,[a] Alison Cheng,[a] Nathanael Joshua Bangayan,[a] Emma Barnard,[a] Emily Curd,[a,b] Noah Craft,[c] Huiying Li[a,d]

Department of Molecular and Medical Pharmacology, Crump Institute for Molecular Imaging, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA[a]; Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, California, USA[b]; Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, USA[c]; UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, California, USA[d]

***Propionibacterium acnes* is a major skin commensal and is associated with acne vulgaris, the most common skin disease. Here we report the draft genome sequences of two *P. acnes* strains, the type strain ATCC6919 and an antibiotic-resistant strain, HL411PA1.**

Received 3 July 2014 **Accepted** 23 July 2014 **Published** 14 August 2014

Citation Liu J, Cheng A, Bangayan NJ, Barnard E, Curd E, Craft N, Li H. 2014. Draft genome sequences of *Propionibacterium acnes* type strain ATCC6919 and antibiotic-resistant strain HL411PA1. Genome Announc. 2(4):e00740-14. doi:10.10.1128/genomeA.00740-14.

Copyright © 2014 Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to Huiying Li, huiying@mednet.ucla.edu.

Propionibacterium acnes is a bacterial species commonly found on human skin. It has been linked to acne pathogenesis. The *P. acnes* type strain ATCC6919 (also known as NCTC737 and DSM1897) was isolated by A. F. Hayden from facial acne (1, 2) and has been a widely used lab strain since the early characterization of *P. acnes* in the 1960s (2–4). As no genomic information was available for ATCC6919, we sequenced this type strain to better understand its documented properties and its virulence potential.

Antibiotic-resistant *P. acnes* strains have been emerging since the 1970s after antibiotics were introduced as one of the main treatments for acne. We isolated a *P. acnes* strain, HL411PA1, from the nose skin of an acne patient who reported no previous antibiotic treatment for acne. However, HL411PA1 is resistant to the major antibiotics used in acne treatment, including tetracycline (MIC 48 μg/mL), erythromycin (MIC >256 μg/mL), and clindamycin (MIC >256 μg/mL), but not minocycline (MIC 0.5 μg/mL). This finding supports previous evidence that antibiotic-resistant strains can be transmitted between individuals (5, 6). We sequenced the genome of HL411PA1 to determine whether this antibiotic-resistant strain encodes novel virulence factors.

The genomes of ATCC6919 and HL411PA1 were sequenced using Illumina MiSeq. The sequence reads were pair-ended with 250 nucleotides per read. Both genomes were assembled using MIRA v3.2.1 (7) with manual inspections using Consed (8). Gene annotation was performed using GeneMark HMM (9) and Glimmer 3.02 (10). Transfer RNAs (tRNAs) and transfer messenger RNAs (tmRNAs) were predicted using Aragorn (11). The ATCC6919 genome is 2,602,215 bp long (260 contigs, N_{50} of 50,832) with 60.1% G+C content and 2,620 predicted open reading frames (ORFs). The sequencing coverage is 72×. The HL411PA1 genome is 2,497,951 bp long (55 contigs, N_{50} of 174,713) with 60.0% G+C content and 2,369 predicted ORFs. The sequencing coverage is 118×.

The genomes of ATCC6919 and HL411PA1 are highly similar to previously sequenced *P. acnes* genomes (6). ATCC6919 and HL411PA1 both cluster with strains from clade IA-1 (12). Based on the 16S rRNA sequences, ATCC6919 belongs to ribotype (RT) 1, a strain type that is prevalent in both acne patients and healthy individuals (12). HL411PA1 belongs to RT5, a strain type associated with acne (12). However, the genome of HL411PA1 is atypical for RT5 in that it lacks any of the three loci found in RT5 strains (6, 12). These loci likely originated from mobile genetic elements and encode putative virulence genes (13).

As ATCC6919 has been widely used in *P. acnes*-related studies, its sequence will provide useful genomic context and insight for both its documented and yet to be discovered properties. The genome of HL411PA1 may contribute to future investigations of antibiotic-resistant *P. acnes* strains in disease pathogenesis.

Nucleotide sequence accession numbers. The whole-genome shotgun projects for ATCC6919 and HL411PA1 were deposited at DDBJ/EMBL/GenBank under the accession numbers JNHS00000000 and JNHT00000000, respectively.

ACKNOWLEDGMENTS

We thank the UCLA GenoSeq core for sequencing the two genomes. All protocols and consent forms of this study were approved by both the UCLA and Los Angeles Biomedical Research Institute IRBs.

This project was funded by the NIH grant R01GM099530 from NIGMS. E.C. was supported by the DOE grant DE-FC03-02ER63421.

REFERENCES

1. **Douglas HC, Gunter SE.** 1946. The taxonomic position of corynebacterium acnes. J. Bacteriol. **52:**15–23.
2. **Zierdt CH, Webster C, Rude WS.** 1968. Study of the anaerobic corynebacteria. Int. J. Syst. Bacteriol. **18:**33–47. http://dx.doi.org/10.1099/00207713-18-1-33.
3. **Voss JG.** 1970. Differentiation of two groups of Corynebacterium acnes. J. Bacteriol. **101:**392–397.
4. **Pulverer G, Ko HL.** 1973. Fermentative and serological studies on Propionibacterium acnes. Appl. Microbiol. **25:**222–229.
5. **Ross JI, Snelling AM, Carnegie E, Coates P, Cunliffe WJ, Bettoli V, Tosti G, Katsambas A, Peréz Del Pulgar JIG, Rollman O, Török L, Eady EA, Cove JH.** 2003. Antibiotic-resistant acne: lessons from Europe. Br. J. Dermatol. **148:**467–478. http://dx.doi.org/10.1046/j.1365-2133.2003.05067.x.
6. **Tomida S, Nguyen L, Chiu B-H, Liu J, Sodergren E, Weinstock GM, Li**

H. 2013. Pan-genome and comparative genome analyses of Propionibacterium acnes reveal its genomic diversity in the healthy and diseased human skin microbiome. mBio **4:**e00003–13. http://dx.doi.org/10.1128/mBio.00003-13.

7. **Chevreux B, Wetter T, Suhai S.** 1999. Genome sequence assembly using trace signals and additional sequence Information, p 45–56. *In* Computer science and biology. Proceedings of the German Conference on Bioinformatics, GCB '99. GCB, Hannover, Germany.
8. **Gordon D, Abajian C, Green P.** 1998. Consed: A graphical tool for sequence finishing. Genome Res. **8:**195–202.
9. **Lukashin AV, Borodovsky M.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26:**1107–1115. doi: http://dx.doi.org/10.1093/nar/26.4.1107.
10. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673–679. http://dx.doi.org/10.1093/bioinformatics/btm009.
11. **Laslett D, Canback B.** 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. **32:**11–16. http://dx.doi.org/10.1093/nar/gkh152.
12. **Fitz-Gibbon S, Tomida S, Chiu B-H, Nguyen L, Du C, Liu M, Elashoff D, Erfe MC, Loncaric A, Kim J, Modlin RL, Miller JF, Sodergren E, Craft N, Weinstock GM, Li H.** 2013. Propionibacterium acnes strain populations in the human skin microbiome associated with acne. J. Invest. Dermatol. **133:**2152–2160. http://dx.doi.org/10.1038/jid.2013.21.
13. **Kasimatis G, Fitz-Gibbon S, Tomida S, Wong M, Li H.** 2013. Analysis of complete genomes of Propionibacterium acnes reveals a novel plasmid and increased pseudogenes in an acne associated strain. Biomed Res. Int. 2013:918320. http://dx.doi.org/10.1155/2013/918320.

References

1. **Cho I**, **Blaser MJ**. 2012. The human microbiome: at the interface of health and disease. Nat Rev Genet **13**:260–270.

2. **Pflughoeft KJ**, **Versalovic J**. 2012. Human microbiome in health and disease. Annu Rev Pathol **7**:99–122.

3. **Hsiao EY**, **McBride SW**, **Hsien S**, **Sharon G**, **Hyde ER**, **McCue T**, **Codelli JA**, **Chow J**, **Reisman SE**, **Petrosino JF**, **Patterson PH**, **Mazmanian SK**. 2013. Microbiota Modulate Behavioral and Physiological Abnormalities Associated with Neurodevelopmental Disorders. Cell **155**:1451–1463.

4. **De Vrese M**, **Schrezenmeir J**. 2008. Probiotics, Prebiotics, and Synbiotics, p. 1–66. *In* Advances in Biochemical Engineering/Biotechnology.

5. **Gao Z**, **Guo B**, **Gao R**, **Zhu Q**, **Wu W**, **Qin H**. 2015. Probiotics modify human intestinal mucosa-associated microbiota in patients with colorectal cancer. Mol Med Rep.

6. **Rodrigues DM**, **Sousa AJ**, **Johnson-Henry KC**, **Sherman PM**, **Gareau MG**. 2012. Probiotics Are Effective for the Prevention and Treatment of Citrobacter rodentium-Induced Colitis in Mice. J Infect Dis **206**:99–109.

7. **Martin F-PJ**, **Wang Y**, **Sprenger N**, **Yap IKS**, **Lundstedt T**, **Lek P**, **Rezzi S**, **Ramadan Z**, **van Bladeren P**, **Fay LB**, **Kochhar S**, **Lindon JC**, **Holmes E**, **Nicholson JK**. 2008. Probiotic modulation of symbiotic gut microbial-host metabolic interactions in a humanized microbiome mouse model. Mol Syst Biol **4**:157.

8. **Kankainen M**, **Paulin L**, **Tynkkynen S**, **von Ossowski I**, **Reunanen J**, **Partanen P**, **Satokari R**, **Vesterlund S**, **Hendrickx AP a**, **Lebeer S**, **De Keersmaecker SCJ**, **Vanderleyden J**, **Hämäläinen T**, **Laukkanen S**, **Salovuori N**, **Ritari J**, **Alatalo E**, **Korpela R**, **Mattila-Sandholm T**, **Lassig A**, **Hatakka K**, **Kinnunen KT**, **Karjalainen H**, **Saxelin M**, **Laakso K**, **Surakka A**, **Palva A**, **Salusjärvi T**, **Auvinen P**, **de Vos WM**. 2009. Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pili containing a human- mucus binding protein. Proc Natl Acad Sci U S A **106**:17193–17198.

9. **Salminen S**, **Nurmi J**, **Gueimonde M**. 2005. The genomics of probiotic intestinal microorganisms. Genome Biol **6**:225.

10. **Buhnik-Rosenblau K**, **Matsko-Efimov V**, **Jung M**, **Shin H**, **Danin-Poleg Y**, **Kashi Y**. 2012. Indication for Co-evolution of Lactobacillus johnsonii with its hosts. BMC Microbiol **12**:149.

11. **Buhnik-Rosenblau K**, **Danin-Poleg Y**, **Kashi Y**. 2011. Predominant Effect of Host Genetics on Levels of Lactobacillus johnsonii Bacteria in the Mouse Gut. Appl Environ Microbiol **77**:6531–6538.

12. **Grazul H**, **Kanda LL**, **Gondek D**. 2016. Impact of probiotic supplements on microbiome diversity following antibiotic treatment of mice. Gut Microbes **0976**:1–14.

13. **Preidis GA**, **Saulnier DM**, **Blutt SE**, **Mistretta T-A**, **Riehle KP**, **Major AM**, **Venable SF**, **Finegold MJ**, **Petrosino JF**, **Conner ME**, **Versalovic J**. 2012. Probiotics stimulate enterocyte migration and microbial diversity in the neonatal mouse intestine. FASEB J **26**:1960–9.

14. **Sattler VA**, **Bayer K**, **Schatzmayr G**, **Haslberger AG**, **Klose V**. 2015. Impact of a probiotic, inulin, or their combination on the piglets' microbiota at different intestinal locations. Benef Microbes **6**:473–483.

15. **Humen MA**, **De Antoni GL**, **Benyacoub J**, **Costas ME**, **Cardozo MI**, **Kozubsky L**, **Saudan K-Y**, **Boenzli-Bruand A**, **Blum S**, **Schiffrin EJ**, **Pérez PF**. 2005. Lactobacillus johnsonii La1 antagonizes Giardia intestinalis in vivo. Infect Immun **73**:1265–9.

16. **Gotteland M**, **Cruchet S**. 2003. Suppressive effect of frequent ingestion of Lactobacillus johnsonii La1 on Helicobacter pylori colonization in asymptomatic volunteers. J Antimicrob Chemother **51**:1317–9.

17. **Isobe H**, **Nishiyama A**, **Takano T**, **Higuchi W**, **Nakagawa S**, **Taneike I**, **Fukushima Y**, **Yamamoto T**. 2012. Reduction of overall Helicobacter pylori colonization levels in the stomach of Mongolian gerbil by Lactobacillus johnsonii La1 (LC1) and its in vitro activities against H. pylori motility and adherence. Biosci Biotechnol Biochem **76**:850–2.

18. **Vidal K**, **Donnet-Hughes A**, **Granato D**. 2002. Lipoteichoic acids from Lactobacillus johnsonii strain La1 and Lactobacillus acidophilus strain La10 antagonize the responsiveness of human intestinal epithelial HT29 cells to lipopolysaccharide and gram-negative bacteria. Infect Immun **70**:2057–64.

19. **Sgouras DN**, **Panayotopoulou EG**, **Martinez-Gonzalez B**, **Petraki K**, **Michopoulos S**,

Mentis A. 2005. Lactobacillus johnsonii La1 attenuates Helicobacter pylori-associated gastritis and reduces levels of proinflammatory chemokines in C57BL/6 mice. Clin Diagn Lab Immunol **12**:1378–86.

20. **Toller IM**, **Neelsen KJ**, **Steger M**, **Hartung ML**, **Hottiger MO**, **Stucki M**, **Kalali B**, **Gerhard M**, **Sartori AA**, **Lopes M**, **Müller A**, **Muller A**. 2011. Carcinogenic bacterial pathogen Helicobacter pylori triggers DNA double-strand breaks and a DNA damage response in its host cells. Proc Natl Acad Sci U S A **108**:14944–14949.

21. **Arthur JC**, **Perez-chanona E**, **Mühlbauer M**, **Tomkovich S**, **Uronis JM**, **Fan T**, **Campbell BJ**, **Abujamel T**, **Dogan B**, **Rogers AB**, **Rhodes JM**, **Stintzi A**, **Simpson KW**, **Hansen JJ**, **Keku TO**, **Fodor A a**, **Jobin C**. 2012. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science **338**:120–3.

22. **Yamamoto ML**, **Maier I**, **Dang AT**, **Berry D**, **Liu J**, **Ruegger PM**, **Yang J-I**, **Soto P a**, **Presley LL**, **Reliene R**, **Westbrook AM**, **Wei B**, **Loy A**, **Chang C**, **Braun J**, **Borneman J**, **Schiestl RH**. 2013. Intestinal Bacteria Modify Lymphoma Incidence and Latency by Affecting Systemic Inflammatory State, Oxidative Stress, and Leukocyte Genotoxicity. Cancer Res **73**:4222–4232.

23. **Touchefeu Y**, **Montassier E**, **Nieman K**, **Gastinne T**, **Potel G**, **Bruley Des Varannes S**, **Le Vacon F**, **De La Cocheti??re MF**, **de La Cochetière MF**. 2014. Systematic review: the role of the gut microbiota in chemotherapy- or radiation-induced gastrointestinal mucositis - current evidence and potential clinical applications. Aliment Pharmacol Ther **40**:409–21.

24. **Manichanh C**, **Varela E**, **Martinez C**, **Antolin M**, **Llopis M**, **Doré J**, **Giralt J**, **Guarner F**, **Malagelada J-R**. 2008. The gut microbiota predispose to the pathophysiology of acute postradiotherapy diarrhea. Am J Gastroenterol **103**:1754–61.

25. **Maier I**, **Berry DM**, **Schiestl RH**. 2014. Intestinal microbiota reduces genotoxic endpoints induced by high-energy protons. Radiat Res **181**:45–53.

26. **Fujiwara D**, **Wei B**, **Presley LL**, **Brewer S**, **McPherson M**, **Lewinski MA**, **Borneman J**, **Braun J**. 2008. Systemic control of plasmacytoid dendritic cells by CD8+ T cells and commensal microbiota. J Immunol **180**:5843–5852.

27. **Rakoff-Nahoum S**, **Paglino J**, **Eslami-Varzaneh F**, **Edberg S**, **Medzhitov R**. 2004. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. Cell **118**:229–41.

28. **Presley LL**, **Wei B**, **Braun J**, **Borneman J**. 2010. Bacteria associated with immunoregulatory cells in mice. Appl Environ Microbiol **76**:936–41.

29. **Ruegger PM**, **Clark RT**, **Weger JR**, **Braun J**, **Borneman J**. 2014. Improved resolution of bacteria by high throughput sequence analysis of the rRNA internal transcribed spacer. J Microbiol Methods **105**:82–7.

30. **Caporaso JG**, **Kuczynski J**, **Stombaugh J**, **Bittinger K**, **Bushman FD**, **Costello EK**, **Fierer N**, **Peña AG**, **Goodrich JK**, **Gordon JI**, **Huttley GA**, **Kelley ST**, **Knights D**, **Koenig JE**, **Ley RE**, **Lozupone CA**, **McDonald D**, **Muegge BD**, **Pirrung M**, **Reeder J**, **Sevinsky JR**, **Turnbaugh PJ**, **Walters WA**, **Widmann J**, **Yatsunenko T**, **Zaneveld J**, **Knight R**. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods **7**:335–336.

31. **Edgar RC**. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26**:2460–1.

32. **DeSantis TZ**, **Hugenholtz P**, **Larsen N**, **Rojas M**, **Brodie EL**, **Keller K**, **Huber T**, **Dalevi D**, **Hu P**, **Andersen GL**. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol **72**:5069–72.

33. **Nam Y-D**, **Kim HJ**, **Seo J-G**, **Kang SW**, **Bae J-W**. 2013. Impact of pelvic radiotherapy on gut microbiota of gynecological cancer patients revealed by massive pyrosequencing. PLoS One **8**:e82659.

34. **Kim YS**, **Kim J**, **Park S-J**. 2015. High-throughput 16S rRNA gene sequencing reveals alterations of mouse intestinal microbiota after radiotherapy. Anaerobe **33**:1–7.

35. **Wang A**, **Ling Z**, **Yang Z**, **Kiela PR**, **Wang T**, **Wang C**, **Cao L**, **Geng F**, **Shen M**, **Ran X**, **Su Y**, **Cheng T**, **Wang J**. 2015. Gut microbial dysbiosis may predict diarrhea and fatigue in patients undergoing pelvic cancer radiotherapy: a pilot study. PLoS One **10**:e0126312.

36. **Vetizou M**, **Pitt JM**, **Daillere R**, **Lepage P**, **Waldschmitt N**, **Flament C**, **Rusakiewicz S**, **Routy B**, **Roberti MP**, **Duong CPM**, **Poirier-Colame V**, **Roux A**, **Becharef S**, **Formenti S**, **Golden E**, **Cording S**, **Eberl G**, **Schlitzer A**, **Ginhoux F**, **Mani S**, **Yamazaki T**, **Jacquelot N**, **Enot DP**, **Berard M**, **Nigou J**, **Opolon P**, **Eggermont A**, **Woerther P-L**, **Chachaty E**, **Chaput N**, **Robert C**, **Mateus C**, **Kroemer G**, **Raoult D**, **Boneca IG**, **Carbonnel F**, **Chamaillard M**, **Zitvogel L**. 2015. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. Science (80-)

350:1079–1084.

37. **Sivan A**, **Corrales L**, **Hubert N**, **Williams JB**, **Aquino-Michaels K**, **Earley ZM**, **Benyamin FW**, **Man Lei Y**, **Jabri B**, **Alegre M-L**, **Chang EB**, **Gajewski TF**. 2015. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. Science (80-) **350**:1084–1089.

38. **Ciorba MA**, **Riehl TE**, **Rao MS**, **Moon C**, **Ee X**, **Nava GM**, **Walker MR**, **Marinshaw JM**, **Stappenbeck TS**, **Stenson WF**. 2012. Lactobacillus probiotic protects intestinal epithelium from radiation injury in a TLR-2/cyclo-oxygenase-2-dependent manner. Gut **61**:829–38.

39. **Bernet-Camard MF**, **Liévin V**, **Brassart D**, **Neeser JR**, **Servin AL**, **Hudault S**. 1997. The human Lactobacillus acidophilus strain LA1 secretes a nonbacteriocin antibacterial substance(s) active in vitro and in vivo. Appl Environ Microbiol **63**:2747–53.

40. **Haller D**, **Bode C**, **Hammes WP**, **Pfeifer AM**, **Schiffrin EJ**, **Blum S**. 2000. Non-pathogenic bacteria elicit a differential cytokine response by intestinal epithelial cell/leucocyte co-cultures. Gut **47**:79–87.

41. **Guinane CM**, **Kent RM**, **Norberg S**, **Hill C**, **Fitzgerald GF**, **Stanton C**, **Ross RP**. 2011. Host specific diversity in Lactobacillus johnsonii as evidenced by a major chromosomal inversion and phage resistance mechanisms. PLoS One **6**:e18740.

42. **Granato D**, **Bergonzelli GE**, **Pridmore RD**, **Marvin L**, **Rouvet M**, **Corthésy-Theulaz IE**. 2004. Cell surface-associated elongation factor Tu mediates the attachment of Lactobacillus johnsonii NCC533 (La1) to human intestinal cells and mucins. Infect Immun **72**:2160–9.

43. **Denou E**, **Pridmore RD**, **Berger B**, **Panoff JM**, **Arigoni F**, **Brüssow H**. 2008. Identification of genes associated with the long-gut-persistence phenotype of the probiotic Lactobacillus johnsonii strain NCC533 using a combination of genomics and transcriptome analysis. J Bacteriol **190**:3161–3168.

44. **Bergonzelli GE**, **Granato D**, **Pridmore RD**, **Marvin-Guy LF**, **Donnicola D**, **Corthesy-Theulaz IE**. 2006. GroEL of Lactobacillus johnsonii La1 (NCC 533) Is Cell Surface Associated: Potential Role in Interactions with the Host and the Gastric Pathogen Helicobacter pylori. Infect Immun **74**:425–434.

45. **Chevreux B**, **Wetter T**, **Suhai S**. 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information, p. 45–56. *In* Computer Science and Biology: Proceedings of the German Conference on Bioinformatics.

46. **Bonfield JK**, **Whitwham A**. 2010. Gap5--editing the billion fragment sequence assembly. Bioinformatics **26**:1699–1703.

47. **Silva GG**, **Dutilh BE**, **Matthews T**, **Elkins K**, **Schmieder R**, **Dinsdale EA**, **Edwards RA**. 2013. Combining de novo and reference-guided assembly with scaffold_builder. Source Code Biol Med **8**:23.

48. **Otto TD**, **Dillon GP**, **Degrave WS**, **Berriman M**. 2011. RATT: Rapid Annotation Transfer Tool. Nucleic Acids Res **39**:e57–e57.

49. **Lukashin a V**, **Borodovsky M**. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res **26**:1107–15.

50. **Delcher a L**, **Harmon D**, **Kasif S**, **White O**, **Salzberg SL**. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res **27**:4636–4641.

51. **Rutherford K**, **Parkhill J**, **Crook J**, **Horsnell T**, **Rice P**, **Rajandream M a**, **Barrell B**. 2000. Artemis: sequence visualization and annotation. Bioinformatics **16**:944–945.

52. **Tomida S**, **Nguyen L**, **Chiu B-H**, **Liu J**, **Sodergren E**, **Weinstock GM**, **Li H**. 2013. Pan-genome and comparative genome analyses of propionibacterium acnes reveal its genomic diversity in the healthy and diseased human skin microbiome. MBio **4**:e00003–13.

53. **Tamura K**, **Peterson D**, **Peterson N**, **Stecher G**, **Nei M**, **Kumar S**. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28**:2731–9.

54. **Edgar RC**. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**:1792–7.

55. **Berger B**, **Pridmore RD**, **Barretto C**, **Delmas-Julien F**, **Schreiber K**, **Arigoni F**, **Brüssow H**. 2007. Similarity and differences in the Lactobacillus acidophilus group identified by polyphasic analysis and comparative genomics. J Bacteriol **189**:1311–1321.

56. **Sun Z**, **Harris HMB**, **McCann A**, **Guo C**, **Argimón S**, **Zhang W**, **Yang X**, **Jeffery IB**, **Cooney JC**, **Kagawa TF**, **Liu W**, **Song Y**, **Salvetti E**, **Wrobel A**, **Rasinkangas P**, **Parkhill J**, **Rea MC**, **O'Sullivan O**, **Ritari J**, **Douillard FP**, **Paul Ross R**, **Yang R**, **Briner AE**, **Felis GE**, **de Vos WM**, **Barrangou R**, **Klaenhammer TR**, **Caufield PW**, **Cui Y**, **Zhang H**, **O'Toole PW**. 2015. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. Nat Commun **6**:8322.

57. **von Schillde M-A**, **Hörmannsperger G**, **Weiher M**, **Alpert C-A**, **Hahne H**, **Bäuerl C**, **van Huynegem K**, **Steidler L**, **Hrncir T**, **Pérez-Martínez G**, **Kuster B**, **Haller D**. 2012. Lactocepin secreted by Lactobacillus exerts anti-inflammatory effects by selectively degrading proinflammatory chemokines. Cell Host Microbe **11**:387–96.

58. **Dieleman LA**, **Goerres MS**, **Arends A**, **Sprengers D**, **Torrice C**, **Hoentjen F**, **Grenther WB**, **Sartor RB**. 2003. Lactobacillus GG prevents recurrence of colitis in HLA-B27 transgenic rats after antibiotic treatment. Gut **52**:370–6.

59. **Justino PFC**, **Melo LFM**, **Nogueira AF**, **Morais CM**, **Mendes WO**, **Franco AX**, **Souza EP**, **Ribeiro RA**, **Souza MHLP**, **Soares PMG**. 2015. Regulatory role of Lactobacillus acidophilus on inflammation and gastric dysmotility in intestinal mucositis induced by 5-fluorouracil in mice. Cancer Chemother Pharmacol **75**:559–567.

60. **Perry A**, **Lambert P**. 2011. Propionibacterium acnes: infection beyond the skin. Expert Rev Anti Infect Ther **9**:1149–56.

61. **Fitz-Gibbon S**, **Tomida S**, **Chiu B-H**, **Nguyen L**, **Du C**, **Liu M**, **Elashoff D**, **Erfe MC**, **Loncaric A**, **Kim J**, **Modlin RL**, **Miller JF**, **Sodergren E**, **Craft N**, **Weinstock GM**, **Li H**. 2013. Propionibacterium acnes strain populations in the human skin microbiome associated with acne. J Invest Dermatol **133**:2152–60.

62. **Kasimatis G**, **Fitz-Gibbon S**, **Tomida S**, **Wong M**, **Li H**. 2013. Analysis of complete genomes of Propionibacterium acnes reveals a novel plasmid and increased pseudogenes in an acne associated strain. Biomed Res Int **2013**:918320.

63. **Liu J**, **Yan R**, **Zhong Q**, **Ngo S**, **Bangayan NJ**, **Nguyen L**, **Lui T**, **Liu M**, **Erfe MC**, **Craft N**, **Tomida S**, **Li H**. 2015. The diversity and host interactions of Propionibacterium acnes bacteriophages on human skin. ISME J **9**:2078–93.

64. **Brown TL**, **Petrovski S**, **Dyson ZA**, **Seviour R**, **Tucci J**. 2016. The Formulation of Bacteriophage in a Semi Solid Preparation for Control of Propionibacterium acnes

Growth. PLoS One **11**:e0151184.

65. **Marinelli LJ**, **Fitz-gibbon S**, **Hayes C**, **Bowman C**, **Inkeles M**, **Loncaric A**, **Russell DA**. 2012. Propionibacterium acnes Bacteriophages Display Limited Genetic Diversity and Broad Killing Activity against Bacterial Skin Isolates. MBio **3**:e00279–12.

66. **Dréno B**, **Poli F**, **Pawin H**, **Beylot C**, **Faure M**, **Chivot M**, **Auffret N**, **Moyse D**, **Ballanger F**, **Revuz J**. 2011. Development and evaluation of a Global Acne Severity Scale (GEA Scale) suitable for France and Europe. J Eur Acad Dermatol Venereol **25**:43–8.

67. **Methé BA**, **Nelson KE**, **Pop M**, **Creasy HH**, **Giglio MG**, **Huttenhower C**, **Gevers D**, **Petrosino JF**, **Abubucker S**, **Badger JH**, **Chinwalla AT**, **Earl AM**, **FitzGerald MG**, **Fulton RS**, **Hallsworth-Pepin K**, **Lobos EA**, **Madupu R**, **Magrini V**, **Martin JC**, **Mitreva M**, **Muzny DM**, **Sodergren EJ**, **Versalovic J**, **Wollam AM**, **Worley KC**, **Wortman JR**, **Young SK**, **Zeng Q**, **Aagaard KM**, **Abolude OO**, **Allen-Vercoe E**, **Alm EJ**, **Alvarado L**, **Andersen GL**, **Anderson S**, **Appelbaum E**, **Arachchi HM**, **Armitage G**, **Arze CA**, **Ayvaz T**, **Baker CC**, **Begg L**, **Belachew T**, **Bhonagiri V**, **Bihan M**, **Blaser MJ**, **Bloom T**, **Bonazzi VR**, **Brooks P**, **Buck GA**, **Buhay CJ**, **Busam DA**, **Campbell JL**, **Canon SR**, **Cantarel BL**, **Chain PS**, **Chen I-MA**, **Chen L**, **Chhibba S**, **Chu K**, **Ciulla DM**, **Clemente JC**, **Clifton SW**, **Conlan S**, **Crabtree J**, **Cutting MA**, **Davidovics NJ**, **Davis CC**, **DeSantis TZ**, **Deal C**, **Delehaunty KD**, **Dewhirst FE**, **Deych E**, **Ding Y**, **Dooling DJ**, **Dugan SP**, **Michael Dunne W**, **Scott Durkin A**, **Edgar RC**, **Erlich RL**, **Farmer CN**, **Farrell RM**, **Faust K**, **Feldgarden M**, **Felix VM**, **Fisher S**, **Fodor AA**, **Forney L**, **Foster L**, **Di Francesco V**, **Friedman J**, **Friedrich DC**, **Fronick CC**, **Fulton LL**, **Gao H**, **Garcia N**, **Giannoukos G**, **Giblin C**, **Giovanni MY**, **Goldberg JM**, **Goll J**, **Gonzalez A**, **Griggs A**, **Gujja S**, **Haas BJ**, **Hamilton HA**, **Harris EL**, **Hepburn TA**, **Herter B**, **Hoffmann DE**, **Holder ME**, **Howarth C**, **Huang KH**, **Huse SM**, **Izard J**, **Jansson JK**, **Jiang H**, **Jordan C**, **Joshi V**, **Katancik JA**, **Keitel WA**, **Kelley ST**, **Kells C**, **Kinder-Haake S**, **King NB**, **Knight R**, **Knights D**, **Kong HH**, **Koren O**, **Koren S**, **Kota KC**, **Kovar CL**, **Kyrpides NC**, **La Rosa PS**, **Lee SL**, **Lemon KP**, **Lennon N**, **Lewis CM**, **Lewis L**, **Ley RE**, **Li K**, **Liolios K**, **Liu B**, **Liu Y**, **Lo C-C**, **Lozupone CA**, **Dwayne Lunsford R**, **Madden T**, **Mahurkar AA**, **Mannon PJ**, **Mardis ER**, **Markowitz VM**, **Mavrommatis K**, **McCorrison JM**, **McDonald D**, **McEwen J**, **McGuire AL**, **McInnes P**, **Mehta T**, **Mihindukulasuriya KA**, **Miller JR**, **Minx PJ**, **Newsham I**, **Nusbaum C**, **O'Laughlin M**, **Orvis J**, **Pagani I**, **Palaniappan K**, **Patel SM**, **Pearson M**, **Peterson J**, **Podar M**, **Pohl C**, **Pollard KS**, **Priest ME**, **Proctor LM**, **Qin X**, **Raes J**, **Ravel J**, **Reid JG**, **Rho M**, **Rhodes R**, **Riehle KP**, **Rivera MC**, **Rodriguez-Mueller B**, **Rogers Y-H**, **Ross MC**, **Russ C**, **Sanka RK**, **Sankar P**, **Fah Sathirapongsasuti J**, **Schloss JA**, **Schloss PD**, **Schmidt TM**, **Scholz M**, **Schriml L**, **Schubert AM**, **Segata N**, **Segre JA**, **Shannon WD**, **Sharp RR**, **Sharpton TJ**, **Shenoy N**, **Sheth NU**, **Simone GA**, **Singh I**, **Smillie CS**, **Sobel JD**, **Sommer DD**, **Spicer P**, **Sutton GG**, **Sykes SM**, **Tabbaa DG**, **Thiagarajan M**, **Tomlinson CM**, **Torralba M**, **Treangen TJ**, **Truty RM**, **Vishnivetskaya TA**, **Walker J**, **Wang L**, **Wang Z**, **Ward D V.**, **Warren W**, **Watson**

MA, **Wellington C**, **Wetterstrand KA**, **White JR**, **Wilczek-Boney K**, **Qing Wu Y**, **Wylie KM**, **Wylie T**, **Yandava C**, **Ye L**, **Ye Y**, **Yooseph S**, **Youmans BP**, **Zhang L**, **Zhou Y**, **Zhu Y**, **Zoloth L**, **Zucker JD**, **Birren BW**, **Gibbs RA**, **Highlander SK**, **Weinstock GM**, **Wilson RK**, **White O**. 2012. A framework for human microbiome research. Nature **486**:215–221.

68. **Kurtz S**, **Phillippy A**, **Delcher AL**, **Smoot M**, **Shumway M**, **Antonescu C**, **Salzberg SL**. 2004. Versatile and open software for comparing large genomes. Genome Biol **5**:R12.

69. **Langmead B**, **Salzberg SL**. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9**:357–359.

70. **Gordon D**, **Abajian C**, **Green P**. 1998. Consed: A Graphical Tool for Sequence Finishing. Genome Res 195–202.